



## Laplace approximation of transition densities posed as Brownian expectations

Markussen, Bo

*Published in:*  
Stochastic Processes and Their Applications

*DOI:*  
[10.1016/j.spa.2008.01.011](https://doi.org/10.1016/j.spa.2008.01.011)

*Publication date:*  
2009

*Document version*  
Early version, also known as pre-print

*Citation for published version (APA):*  
Markussen, B. (2009). Laplace approximation of transition densities posed as Brownian expectations. *Stochastic Processes and Their Applications*, 119(1), 208-231. <https://doi.org/10.1016/j.spa.2008.01.011>

# Laplace approximation of transition densities posed as Brownian expectations

Bo Markussen\*(email: bomar@life.ku.dk)

Department of Basic Sciences and Environment,  
Faculty of Life Sciences, University of Copenhagen  
Thorvaldsensvej 40, DK-1871 Frederiksberg C, Denmark

## Abstract

We construct the Laplace approximation of the Lebesgue density for a discrete partial observation of a multidimensional stochastic differential equation. This approximation may be computed integrating systems of ordinary differential equations. The construction of the Laplace approximation begins with the definition of the point of minimum energy. We show how such a point can be defined in the Cameron-Martin space as a maximum a posteriori estimator of the underlying Brownian motion given an observation of a finite dimensional functional. The definition of the MAP estimator is possible via a renormalization of the densities of piecewise linear approximations of the Brownian motion. Using the renormalized Brownian density the Laplace approximation of the integral over all Brownian paths can be defined. The developed theory provides a method for performing approximate maximum likelihood estimation.

*Key words:* Stochastic differential equation, maximum a posteriori estimation, maximum likelihood estimation, discrete partial observation, renormalized Brownian density, white noise, path integral, Laplace approximation  
*1991 MSC:* 60H10, 60H35, 60H40, 47N30

## 1 Introduction

Functionals of Brownian motion play a role in many branches of science, notably biology, chemistry, physics and mathematical finance, and applications

---

\*Supported by the Danish Research Agency project “Computing Natural Shape”, identification number 56-00-0106.

are emerging in new areas such as computer vision and image analysis. An often investigated model is the class of  $q$ -dimensional diffusion processes  $X$  solving a stochastic differential equation driven by a  $d$ -dimensional Brownian motion  $B$  with Itô drift  $\xi: \mathbb{R}^q \rightarrow \mathbb{R}^q$  and diffusion coefficient  $\sigma: \mathbb{R}^q \rightarrow \mathbb{R}^{q \times d}$ ,

$$dX(t) = \xi(X(t))dt + \sigma(X(t))dB(t), \quad X(0) = x_0. \quad (1)$$

In parametric models a family of coefficients  $(\xi, \sigma) = (\xi_\theta, \sigma_\theta)$  indexed by  $\theta \in \Theta$  is available, and the process  $X$  is assumed to satisfy eq. (1) for some  $\theta_0 \in \Theta$ . Estimation of the true parameter  $\theta_0$  given the observation of  $X$  at prefixed discrete time points has received much attention in the last decade. Since the transition densities for the solution to SDEs rarely are explicitly known various alternatives to a closed form likelihood approach have been developed. Among such procedures are martingale methods [6], simulation techniques [5], numerical solution of the forward Kolmogorov equation [19], and series expansions of the forward Kolmogorov equation [1, 2]. In this paper we provide a new approximation of the Lebesgue density of an  $M$ -dimensional functional of the Brownian motion given on the form

$$\{P_j X(t_j)\}_{j=1, \dots, J} \in \mathbb{R}^M, \quad M = \sum_{j=1}^J q_j, \quad (2)$$

where  $t_1 < \dots < t_J$  and  $P_j \in \mathbb{R}^{q_j \times q}$  are given full rank matrices with  $q_j \leq q$ . The matrices  $P_j$  can be used to pick out the observed components of partially observed multivariate SDEs, e.g. stochastic volatility models [11].

The Itô equation (1) can be written on Stratonovich form replacing  $\xi(x)$  with the Stratonovich drift  $b(x) = \xi(x) - \frac{1}{2} \nabla \sigma(x) \cdot \sigma(x)$ . If  $\phi_x^{s,t}(Z)$  denotes the flow of diffeomorphisms generated by a  $d$ -dimensional semimartingale  $Z$  via the Stratonovich equation

$$\phi_x^{s,t}(Z) = x + \int_s^t b(\phi_x^{s,u}(Z))du + \int_s^t \sigma(\phi_x^{s,u}(Z)) \circ dZ(u), \quad (3)$$

then  $X(t) = \phi_{x_0}^{0,t}(B)$  solves eq. (1). For fixed  $T > 0$  the solution  $X(T)$  varies continuously with the initial position  $x_0$ , i.e. the mapping  $x \mapsto \phi_x^{0,T}(B)$  is a stochastic diffeomorphism of  $\mathbb{R}^q$ . Such mappings are known as warps in the image analysis literature and are e.g. used to align different image modalities [23]. Markussen [14, 15] introduced a renormalization technique to define the maximum a posteriori alignment. In this paper the renormalization is used to derive the Laplace approximation of the Lebesgue density introduced above. Invoking the dependence on the parameter  $\theta$  this provides an approximation of the likelihood function given the discrete partial observation defined in eq. (2).

The idea behind our construction is to rewrite the Lebesgue density in question as an expectation w.r.t. the Brownian motion, and to derive the

Laplace approximation of this expectation. The idea may be illustrated for the observation of  $X$  at time  $T > 0$ : If  $R$  is a standard  $q$ -dimensional Gaussian random variable independent of  $B$ , then  $X(T) + \epsilon R$  converges in probability to  $X(T)$  as  $\epsilon \rightarrow 0$  and the transition density  $p(x_T)$  from  $X(0) = x_0$  to  $X(T) = x_T$  satisfies

$$p(x_T) = \lim_{\epsilon \rightarrow 0} \mathbb{E} \left[ \frac{1}{(2\pi\epsilon^2)^{q/2}} \exp \left( - \frac{|X(T) - x_T|^2}{2\epsilon^2} \right) \right]. \quad (4)$$

Since  $X(T)$  is a functional of the Brownian motion, the transition density is written as an exponential integral over the path space  $\mathcal{C}([0, T]; \mathbb{R}^d)$  w.r.t. the Brownian measure. To define the Laplace approximation we consider the analogy with the exponential integral  $\int_{\mathbb{R}^N} \exp(-E(x)) dx$  of an energy function  $E \in \mathcal{C}^2(\mathbb{R}^N; \mathbb{R})$  on the  $N$ -dimensional Euclidean space. The Laplace approximation is found approximating  $E(x)$  by its second order Taylor expansion around the point of minimum energy, i.e.

$$E(x) \approx E(\hat{x}) + \frac{1}{2}(x - \hat{x})^\top \hat{A}(x - \hat{x}), \quad \hat{x} = \arg \min_{x \in \mathbb{R}^N} E(x),$$

where  $\hat{A}$  is the matrix  $D^2 E(x)$  of second order partial derivatives evaluated at  $x = \hat{x}$ . This leads to the approximation given by a Gaussian integral

$$\int_{\mathbb{R}^N} e^{-E(x)} dx \approx \int_{\mathbb{R}^N} e^{-E(\hat{x}) - \frac{1}{2}(x - \hat{x})^\top \hat{A}(x - \hat{x})} dx = (2\pi)^{N/2} e^{-E(\hat{x})} (\det \hat{A})^{-1/2}.$$

This analogy identifies the steps needed to define the Laplace approximation of eq. (4): We must define a Brownian energy functional, find the minimum energy path  $\hat{B}$  given  $X(T) = x_T$ , and compute the determinant of the second order variational derivative of the energy functional at the point of minimum energy. Below we elaborate further on these issues and describe the organization of the paper.

Seemingly the construction of the Laplace approximation of eq. (4) is prohibited by the non-differentiability of the Brownian paths. To circumvent this obstacle we approximate the Brownian paths by smooth functions. Our construction is done in the Cameron-Martin space  $\mathcal{H}$  of continuous functions  $f$  with  $f(0) = 0$  and weak derivatives  $\partial f \in \mathcal{K} = L^2([0, T]; \mathbb{R}^d)$ ,

$$\mathcal{H} = \left\{ f \in \mathcal{C}([0, T]; \mathbb{R}^d) \mid f(t) = \int_0^t \partial_s f(s) ds, \partial f \in \mathcal{K} \right\}. \quad (5)$$

In Section 3 a Brownian energy functional is defined as a renormalization of minus the log probability density for a piecewise linear approximation of the Brownian motion. We demonstrate how the minimum energy path  $\hat{B}$  can be conceived as the maximum a posteriori (MAP) estimator in a Bayesian framework [4], and the Stratonovich representation eq. (3) turns

out to be appropriate in this context. The interpretation of  $\widehat{B}$  is the most likely path of the Brownian motion given the observation  $X(T) = x_T$ . This path is in unique correspondence with the minimum energy path  $\widehat{X}$  from  $X(0) = x_0$  to  $X(T) = x_T$ , and the specialization of Theorem 5 to the univariate case  $q = d = 1$  gives that  $\widehat{X}$  solves the second order non-linear ordinary differential equation

$$\partial_t^2 \widehat{X}(t) = \frac{\partial_x \sigma(\widehat{X}(t))}{\sigma(\widehat{X}(t))} \left( (\partial_t \widehat{X}(t))^2 - b(\widehat{X}(t))^2 \right) + b(\widehat{X}(t)) \partial_x b(\widehat{X}(t)).$$

In Section 4 we describe the second order variational derivative of the energy functional in the exponential integral eq. (4). This variational derivative measures the curvature of the energy surface around the energy minimum and will be referred to as the curvature operator. We show that the curvature operator belongs to a class of kernel operators with a lattice structure. In Section 5 we derive the Laplace approximation of the Lebesgue density for the discrete partial observation  $Y = \{P_j X(t_j)\}$  described in eq. (2). The main result of the paper stated in Theorem 6 is the approximation

$$p(y) \approx \frac{\exp \left( -\frac{1}{2} \|\partial \widehat{B}\|_{\mathcal{K}}^2 - \frac{1}{2} \text{tr}_{\mathcal{K}} \log(\mathbb{I}_{\mathcal{K}} + \widehat{Q} \widehat{A} \widehat{Q}) \right)}{(2\pi)^{M/2} \sqrt{\det \left( \int_0^T \gamma_{\widehat{B}}(t)^\top \gamma_{\widehat{B}}(t) dt \right)}}.$$

Here the function  $\gamma_{\widehat{B}}(t)$  is defined in Proposition 2,  $\widehat{Q}$  is the projection of  $\mathcal{K}$  on the orthogonal complement of the subspace spanned by  $\gamma_{\widehat{B}}$ , and  $\widehat{A}$  is the non-degenerate part of the curvature operator. The logarithmic trace term can be computed via inversion of the operators  $\mathbb{I}_{\mathcal{K}} + z\widehat{A}$ ,  $z \in [0, 1]$ , and an inversion formula for such operators is stated. These formulae allow the Laplace approximation to be computed integrating systems of ordinary differential equations. In the appendix we explicitly state the systems to be solved in the univariate case  $q = d = 1$ . Similar systems of ODEs apply to multivariate SDEs. Finally, in Section 6 we discuss the developed theory and provide some examples.

## 2 Function spaces and regularity conditions

In this section we provide a set of regularity conditions ensuring the validity of our analysis. But first we list the used function spaces, operators and seminorms. All these quantities are defined relative to a fixed time horizon  $T > 0$  and fixed dimensions  $q, d \in \mathbb{N}$ .

The elements in the spaces  $\mathcal{K}$  and  $\mathcal{H}$  are paired via  $\partial f \in \mathcal{K}$  and  $f \in \mathcal{H}$ , or emphasizing the argument as  $\partial_t f(t)$  and  $f(t)$ , cf. eq. (5). The spaces  $\mathcal{K}$  and  $\mathcal{H}$  are Hilbert spaces equipped with the norms

$$\|f\|_{\mathcal{H}} = \|\partial f\|_{\mathcal{K}} = \left( \int_0^T |\partial_t f(t)|^2 dt \right)^{1/2}, \quad f \in \mathcal{H}, \partial f \in \mathcal{K}.$$

For each  $N \in \mathbb{N}$  let  $\mathcal{H}_N = L_N(\mathbb{R}^{d \times N})$  be the subspace of functions  $f \in \mathcal{H}$  that are piecewise linear on the intervals  $(\frac{(n-1)T}{N}, \frac{nT}{N})$ , and let the operators  $L_N: \mathbb{R}^{d \times N} \rightarrow \mathcal{H}$  and  $PL_N: \mathcal{C}([0, T]; \mathbb{R}^d) \rightarrow \mathcal{H}$  be defined by

$$L_N(\{u_n\}_{n=1, \dots, N})(t) = \left( \left( n - \frac{tN}{T} \right) u_{n-1} + \left( \frac{tN}{T} - n + 1 \right) u_n \right) \Big|_{n=\lceil \frac{tN}{T} \rceil, u_0=0},$$

$$PL_N(f)(t) = \left( \left( n - \frac{tN}{T} \right) f\left(\frac{(n-1)T}{N}\right) + \left( \frac{tN}{T} - n + 1 \right) f\left(\frac{nT}{N}\right) \right) \Big|_{n=\lceil \frac{tN}{T} \rceil}.$$

For  $\alpha = (\alpha_1, \dots, \alpha_q) \in \mathbb{N}_0^q$  the differential operators  $D^\alpha$  are defined by

$$D^\alpha f(x) = \frac{\partial^{|\alpha|} f(x)}{(\partial x_1)^{\alpha_1} \dots (\partial x_q)^{\alpha_q}}, \quad |\alpha| = \alpha_1 + \dots + \alpha_q.$$

For  $k \in \mathbb{N}$  and  $0 < \eta \leq 1$  let  $\mathcal{C}_b^{k, \eta}(\mathbb{R}^q; \mathbb{R}^q)$  be the vector space of  $k$ -times differentiable functions  $f: \mathbb{R}^q \rightarrow \mathbb{R}^q$  with  $\|f\|_{k+\eta} < \infty$ , where

$$\|f\|_{k+\eta} = \sup_{x \in \mathbb{R}^q} \frac{|f(x)|}{1 + |x|} + \sum_{\substack{\alpha \in \mathbb{N}_0^q: \\ 1 \leq |\alpha| \leq k}} \sup_{x \in \mathbb{R}^q} |D^\alpha f(x)| + \sum_{\substack{\alpha \in \mathbb{N}_0^q: \\ |\alpha| = k}} \sup_{\substack{x, y \in \mathbb{R}^q: \\ x \neq y}} \frac{|D^\alpha f(x) - D^\alpha f(y)|}{|x - y|^\eta}.$$

Let  $\mathcal{D}^k$  be the space of continuous functions  $f: \mathbb{R}^q \times [0, T] \rightarrow \mathbb{R}^q$  such that  $f(\cdot, t): \mathbb{R}^q \rightarrow \mathbb{R}^q$  is a  $\mathcal{C}^k$ -diffeomorphism for every  $t \in [0, T]$ . The inverse mapping  $f^\leftarrow \in \mathcal{D}^k$  is defined separately for each  $t$ , i.e.  $f(\cdot, t) \circ f^\leftarrow(\cdot, t)$  is the identity mapping on  $\mathbb{R}^q$ . The space  $\mathcal{D}^k$  is equipped with the topology induced by the seminorms  $\|\cdot\|_{k:r}$  and  $\|\cdot^\leftarrow\|_{k:r}$  for  $r \in \mathbb{N}$ , where

$$\|f\|_{k:r} = \sum_{\alpha \in \mathbb{N}_0^q: |\alpha| \leq k} \sup_{t \in [0, T]} \sup_{x \in \mathbb{R}^q: |x| \leq r} |D^\alpha f(x, t)|.$$

We now state the regularity conditions on the Stratonovich drift  $b: \mathbb{R}^q \rightarrow \mathbb{R}^q$  and the diffusion coefficient  $\sigma = \{\sigma_i\}_{i=1, \dots, d}$  with  $\sigma_i: \mathbb{R}^q \rightarrow \mathbb{R}^q$ .

**(A1)** There exists  $k \geq 2$  and  $0 < \eta \leq 1$  such that  $b \in \mathcal{C}_b^{k, \eta}(\mathbb{R}^q; \mathbb{R}^q)$  and  $\sigma_i \in \mathcal{C}_b^{k+1, \eta}(\mathbb{R}^q; \mathbb{R}^q)$  for every  $i = 1, \dots, d$ .

**(A2)** The matrix  $\sigma(x) \sigma(x)^\top \in \mathbb{R}^{q \times q}$  is invertible for every  $x \in \mathbb{R}^q$ .

**(A3)** We have  $\sup_{x \in \mathbb{R}^q} |b(x)| < \infty$ , and there exists a positive definite matrix  $\Gamma \in \mathbb{R}^{q \times q}$  such that  $\sigma(x) \sigma(x)^\top \geq \Gamma$  for every  $x \in \mathbb{R}^q$ .

Finally, we state theorems from stochastic calculus underlying our construction. The first part of Theorem 2 follows from [13, Theorem 4.6.5] using the relationship between Itô and Stratonovich stochastic differential equations [13, Theorem 3.4.7]. The second part is known as the Wong-Zakai theorem, and is stated as [13, Theorem 5.7.3 and Example 5.7.4]. Theorem 1 is a consequence of the proof of the Wong-Zakai theorem. Theorem 3

follows by [17, Remark 2, page 123]. In the statement of these theorems the operator  $\Phi$  is defined via the solution to eq. (3), i.e.

$$\Phi(Z)(x, t) = \phi_x^{0, t}(Z), \quad Z \text{ semimartingale, } (x, t) \in \mathbb{R}^q \times [0, T].$$

**Theorem 1.** *The elements in the Cameron-Martin space have finite variation and are semimartingales. If (A1) holds true, then the solution operator  $\Phi: \mathcal{H} \rightarrow \mathcal{D}^{k-1}$  is uniquely defined and continuous.*

**Theorem 2.** *If (A1) holds true, then there exists a unique solution  $\Phi(B)$  to eq. (3) with  $Z = B$  being the Brownian motion. This solution constitutes a flow of  $\mathcal{C}^k$ -diffeomorphisms. If  $B_N = \text{PL}_N(B)$  denotes the  $N$ 'th piecewise linear approximation of the Brownian motion, then  $\Phi(B_N)$  converge to  $\Phi(B)$  as elements of  $\mathcal{D}^{k-1}$  as  $N \rightarrow \infty$  in  $L^p$  for every  $p > 1$ , i.e.*

$$\mathbb{E} \left( \|\Phi(B) - \Phi(B_N)\|_{k-1:r}^p + \|\Phi(B)^\leftarrow - \Phi(B_N)^\leftarrow\|_{k-1:r}^p \right) \xrightarrow{N \rightarrow \infty} 0, \quad r \in \mathbb{N}.$$

**Theorem 3.** *If (A1) and (A2) hold true, then the transition probabilities for the diffusion process  $X(t) = \phi_x^{0, t}(B)$  have continuous Lebesgue densities.*

### 3 Renormalized Brownian density and MAP estimation

Suppose that a bounded continuous functional  $\Upsilon: \mathcal{D}^{k-1} \rightarrow \mathbb{R}^M$  of the flow of diffeomorphisms  $\Phi(B)$  is observed. Let  $B_N = \text{PL}_N(B) \in \mathcal{H}_N$  be the  $N$ 'th piecewise linear approximation of the Brownian motion, and let  $R$  be a standard  $M$ -dimensional Gaussian random variable independent of  $B$ . For  $N \in \mathbb{N}$  and  $\epsilon > 0$  we define

$$Y = (\Upsilon \circ \Phi)(B), \quad Y^{(\epsilon)} = (\Upsilon \circ \Phi)(B) + \epsilon R, \quad Y_N^{(\epsilon)} = (\Upsilon \circ \Phi)(B_N) + \epsilon R.$$

The random variable  $R$  is a technical device serving two purposes. Firstly, the MAP estimator  $\hat{B}_N^{(\epsilon)}$  for  $B_N$  given the observation of  $Y_N^{(\epsilon)}$  can be defined maximizing over  $\mathbb{R}^{d \times N}$ . Secondly, the Lebesgue density of the functional  $Y$  can be written similarly as in eq. (4).

**Proposition 1.** *The pair  $(Y_N^{(\epsilon)}, B_N) = (y, f) \in \mathbb{R}^M \times \mathcal{H}_N$  has a density w.r.t.  $\text{Leb}(\mathbb{R}^M) \otimes L_N(\text{Leb}(\mathbb{R}^{d \times N}))$ , namely  $p^{(\epsilon)}(y|f) p_N(f)$  with*

$$p^{(\epsilon)}(y|f) = (2\pi\epsilon^2)^{-M/2} \exp \left( -\frac{|y - (\Upsilon \circ \Phi)(f)|^2}{2\epsilon^2} \right), \quad (y, f) \in \mathbb{R}^M \times \mathcal{H},$$

$$p_N(f) = \prod_{n=1}^N \left( \frac{N}{2\pi T} \right)^{d/2} \exp \left( -\frac{N |f(\frac{nT}{N}) - f(\frac{(n-1)T}{N})|^2}{2T} \right), \quad f \in \mathcal{H}_N. \quad (6)$$

Furthermore, there exists a MAP estimator  $\widehat{B}_N^{(\epsilon)} \in \mathcal{H}_N$  for  $B_N \in \mathcal{H}_N$  given the observation of  $Y_N^{(\epsilon)} \in \mathbb{R}^M$ , i.e.

$$p^{(\epsilon)}(Y_N^{(\epsilon)} | \widehat{B}_N^{(\epsilon)}) p_N(\widehat{B}_N^{(\epsilon)}) = \sup_{f \in \mathbb{R}^{d \times N}} p^{(\epsilon)}(Y_N^{(\epsilon)} | L_N(f)) p_N(L_N(f)). \quad (7)$$

**Proof.** The form of the densities in eq. (6) is evident. To show the existence of  $\widehat{B}_N^{(\epsilon)}$  we choose a sequence  $\{f_k\}_{k \in \mathbb{N}}$  with  $f_k \in \mathbb{R}^{d \times N}$ , which attains the right hand side of eq. (7) as  $k \rightarrow \infty$ . It is clear that the points  $f_k$  are contained in a bounded ball of  $\mathbb{R}^{d \times N}$ . By compactness there exists  $\widehat{f} \in \mathbb{R}^{d \times N}$  and a subsequence  $k'_n$  such that  $\lim_{n \rightarrow \infty} f_{k'_n} = \widehat{f}$ , and we may choose  $\widehat{B}_N^{(\epsilon)} = L_N(\widehat{f})$ .  $\square$

The density  $p_N(f)$  for  $B_N = f \in \mathcal{H}_N$  satisfies

$$\begin{aligned} p_N(f) &= \prod_{n=1}^N \left( \frac{N}{2\pi T} \right)^{d/2} \exp \left( - \frac{N |f(\frac{nT}{N}) - f(\frac{(n-1)T}{N})|^2}{2T} \right) \\ &= \left( \frac{N}{2\pi T} \right)^{Nd/2} \exp \left( - \frac{\|f\|_{\mathcal{H}}^2}{2} \right). \end{aligned} \quad (8)$$

Since the factor  $p_N(0) = (N/2\pi T)^{Nd/2}$  is immaterial for the purpose of MAP estimation, we define the renormalized Brownian density by

$$p_{\text{renorm}}(f) = \frac{p_N(f)}{p_N(0)} = \exp \left( - \frac{\|f\|_{\mathcal{H}}^2}{2} \right), \quad f \in \mathcal{H}_N \subset \mathcal{H}.$$

This definition is independent of  $N$ , and  $p_{\text{renorm}}$  extends continuously to the entire Cameron-Martin space. Using the renormalized Brownian density we define the MAP estimators  $\widehat{B}, \widehat{B}^{(\epsilon)} \in \mathcal{H}$  given the observation of  $Y, Y^{(\epsilon)} \in \mathbb{R}^M$ , respectively, by

$$\widehat{B}^{(\epsilon)} = \arg \max_{f \in \mathcal{H}} p^{(\epsilon)}(Y^{(\epsilon)} | f) p_{\text{renorm}}(f), \quad \widehat{B} = \arg \max_{f \in \mathcal{H}: (\Upsilon \circ \Phi)(f) = Y} p_{\text{renorm}}(f).$$

Existence and uniqueness of  $\widehat{B}$  and  $\widehat{B}^{(\epsilon)}$  is discussed in Section 3.1 assuming a particular form of the functional  $\Upsilon$ . We remark, that since  $P(B \in \mathcal{H}) = 0$  the a priori and the a posteriori distributions are mutually singular. Nevertheless, the validity of  $\widehat{B}$  as a MAP estimator for  $B$  given the observation  $Y$  can be established via the relationships displayed below:

$$\begin{array}{ccccccc} B_N & \xrightarrow{\quad} & Y_N^{(\epsilon)} & \xrightarrow{\text{MAP}} & \widehat{B}_N^{(\epsilon)} & \xrightarrow{\quad} & \Phi(\widehat{B}_N^{(\epsilon)}) \\ N \rightarrow \infty \downarrow \text{a.s.} & & N \rightarrow \infty \downarrow L^p & & N \rightarrow \infty \downarrow \text{prob} & & N \rightarrow \infty \downarrow \text{prob} \\ B & \xrightarrow{\quad} & Y^{(\epsilon)} & \xrightarrow{\quad} & \widehat{B}^{(\epsilon)} & \xrightarrow{\quad} & \Phi(\widehat{B}^{(\epsilon)}) \\ \parallel & & \epsilon \rightarrow 0 \downarrow \text{a.s.} & & \epsilon \rightarrow 0 \downarrow \text{a.s.} & & \epsilon \rightarrow 0 \downarrow \text{a.s.} \\ B & \xrightarrow{\quad} & Y & \xrightarrow{\quad} & \widehat{B} & \xrightarrow{\quad} & \Phi(\widehat{B}) \end{array} \quad (9)$$



The only non-trivial convergence result in the first two columns of diagram (9) is  $Y_N^{(\epsilon)} \rightarrow Y^{(\epsilon)}$  in  $L^p$  as  $N \rightarrow \infty$ , which follows by Theorem 2 and the properties of the functional  $\Upsilon$ . Suppose that the convergence results displayed in the last two columns of diagram (9) have been proven. Then the mapping  $Y \mapsto \hat{B}$  qualifies as a MAP estimator. We may namely from an arbitrarily good approximation  $Y_N^{(\epsilon)}$  of the observation  $Y$  find a standard MAP estimator  $\hat{B}_N^{(\epsilon)}$  of the arbitrarily good approximation  $B_N$  of  $B$ , and this MAP estimator is an arbitrarily good approximation of the estimator  $\hat{B}$ . The same argument qualifies the mapping  $Y \mapsto \Phi(\hat{B})$  as a MAP estimator for the flow of diffeomorphisms  $\Phi(B)$  given the observation  $Y$ .

**Theorem 4.** *Assume condition (A1) and let  $\epsilon > 0$  be given. If the estimator  $\hat{B}^{(\epsilon)}$  is uniquely defined and well-separated, i.e. for every  $\delta > 0$  we have*

$$p^{(\epsilon)}(Y^{(\epsilon)}|\hat{B}^{(\epsilon)}) p_{\text{renorm}}(\hat{B}^{(\epsilon)}) > \sup_{f \in \mathcal{H}: \|f - \hat{B}^{(\epsilon)}\|_{\mathcal{H}} > \delta} p^{(\epsilon)}(Y^{(\epsilon)}|f) p_{\text{renorm}}(f),$$

*then  $\|\hat{B}_N^{(\epsilon)} - \hat{B}^{(\epsilon)}\|_{\mathcal{H}}$ ,  $\|\Phi(\hat{B}_N^{(\epsilon)}) - \Phi(\hat{B}^{(\epsilon)})\|_{k-1:r}$  and  $\|\Phi(\hat{B}_N^{(\epsilon)})^{\leftarrow} - \Phi(\hat{B}^{(\epsilon)})^{\leftarrow}\|_{k-1:r}$  for  $r \in \mathbb{N}$  vanish in probability as  $N \rightarrow \infty$ .*

**Proof.** *Let the random variables  $U_N^{(\epsilon)}$  be defined by*

$$U_N^{(\epsilon)} = \sup_{f \in \mathcal{H}} |p^{(\epsilon)}(Y_N^{(\epsilon)}|f) - p^{(\epsilon)}(Y^{(\epsilon)}|f)|.$$

*Theorem 2 and equicontinuity of the functions  $p^{(\epsilon)}(\cdot|f)$  over  $f \in \mathcal{H}$ , i.e.*

$$\begin{aligned} \sup_{f \in \mathcal{H}} |p^{(\epsilon)}(x|f) - p^{(\epsilon)}(y|f)| &\leq (2\pi\epsilon^2)^{-M/2} \sup_{z \in \mathbb{R}^M} \left\{ e^{-\frac{|x-z|^2}{2\epsilon^2}} - e^{-\frac{|y-z|^2}{2\epsilon^2}} \right\} \\ &= (2\pi\epsilon^2)^{-M/2} \sup_{u \in \mathbb{R}} \left\{ e^{-\frac{(u-|x-y|)^2}{2\epsilon^2}} - e^{-\frac{u^2}{2\epsilon^2}} \right\} \xrightarrow{|x-y| \rightarrow 0} 0, \end{aligned}$$

*imply that  $U_N^{(\epsilon)}$  vanish in probability as  $N \rightarrow \infty$ . Thus, for any subsequence  $N'_n$  of the natural numbers there exists a further subsequence  $N''_n$  such that  $U_{N''_n}^{(\epsilon)}$  vanish almost surely as  $n \rightarrow \infty$ . Employing this together with the bound*

$p_{\text{renorm}}(f) \leq 1$  and the properties of the MAP estimators we have

$$\begin{aligned}
& p^{(\epsilon)}(Y^{(\epsilon)}|\widehat{B}^{(\epsilon)}) p_{\text{renorm}}(\widehat{B}^{(\epsilon)}) \\
&= \lim_{n \rightarrow \infty} p^{(\epsilon)}(Y^{(\epsilon)}|\text{PL}_{N_n''}(\widehat{B}^{(\epsilon)})) p_{\text{renorm}}(\text{PL}_{N_n''}(\widehat{B}^{(\epsilon)})) \\
&\leq \liminf_{n \rightarrow \infty} \left( p^{(\epsilon)}(Y_{N_n''}^{(\epsilon)}|\text{PL}_{N_n''}(\widehat{B}^{(\epsilon)})) + U_{N_n''}^{(\epsilon)} \right) p_{\text{renorm}}(\text{PL}_{N_n''}(\widehat{B}^{(\epsilon)})) \\
&= \liminf_{n \rightarrow \infty} p^{(\epsilon)}(Y_{N_n''}^{(\epsilon)}|\text{PL}_{N_n''}(\widehat{B}^{(\epsilon)})) p_{\text{renorm}}(\text{PL}_{N_n''}(\widehat{B}^{(\epsilon)})) \\
&\leq \liminf_{n \rightarrow \infty} p^{(\epsilon)}(Y_{N_n''}^{(\epsilon)}|\widehat{B}_{N_n''}^{(\epsilon)}) p_{\text{renorm}}(\widehat{B}_{N_n''}^{(\epsilon)}) \\
&\leq \liminf_{n \rightarrow \infty} \left( p^{(\epsilon)}(Y^{(\epsilon)}|\widehat{B}_{N_n''}^{(\epsilon)}) + U_{N_n''}^{(\epsilon)} \right) p_{\text{renorm}}(\widehat{B}_{N_n''}^{(\epsilon)}) \\
&= \liminf_{n \rightarrow \infty} p^{(\epsilon)}(Y^{(\epsilon)}|\widehat{B}_{N_n''}^{(\epsilon)}) p_{\text{renorm}}(\widehat{B}_{N_n''}^{(\epsilon)}) \\
&\leq \limsup_{n \rightarrow \infty} p^{(\epsilon)}(Y^{(\epsilon)}|\widehat{B}_{N_n''}^{(\epsilon)}) p_{\text{renorm}}(\widehat{B}_{N_n''}^{(\epsilon)}) \\
&\leq p^{(\epsilon)}(Y^{(\epsilon)}|\widehat{B}^{(\epsilon)}) p_{\text{renorm}}(\widehat{B}^{(\epsilon)})
\end{aligned}$$

almost surely. Thus, we have

$$p^{(\epsilon)}(Y^{(\epsilon)}|\widehat{B}_{N_n''}^{(\epsilon)}) p_{\text{renorm}}(\widehat{B}_{N_n''}^{(\epsilon)}) \xrightarrow[n \rightarrow \infty]{a.s.} p^{(\epsilon)}(Y^{(\epsilon)}|\widehat{B}^{(\epsilon)}) p_{\text{renorm}}(\widehat{B}^{(\epsilon)}).$$

Since  $\widehat{B}^{(\epsilon)}$  is assumed to be well-separated this implies  $\|\widehat{B}_{N_n''}^{(\epsilon)} - \widehat{B}^{(\epsilon)}\|_{\mathcal{H}} \xrightarrow[n \rightarrow \infty]{a.s.} 0$ , whereby  $\|\Phi(\widehat{B}_{N_n''}^{(\epsilon)}) - \Phi(\widehat{B}^{(\epsilon)})\|_{k-1:r} + \|\Phi(\widehat{B}_{N_n''}^{(\epsilon)})^{\leftarrow} - \Phi(\widehat{B}^{(\epsilon)})^{\leftarrow}\|_{k-1:r} \xrightarrow[n \rightarrow \infty]{a.s.} 0$  follows by Theorem 1. Since this holds for any subsequence  $N_n'$  we have demonstrated convergence in probability as desired.  $\square$

### 3.1 Discretely observed stochastic differential equations

In this section we study the MAP estimators  $\widehat{B}$ ,  $\widehat{B}^{(\epsilon)}$  given the discrete partial observation introduced in Section 1. This analysis will complete the lower right corner of the diagram (9). Let  $x_0 \in \mathbb{R}^q$ ,  $0 = t_0 < t_1 < \dots < t_J = T$ , and full rank matrices  $P_j \in \mathbb{R}^{q_j \times q}$  with  $q_j \leq q$  be fixed, and consider the bounded linear functional  $\Upsilon: \mathcal{D}^{k-1} \rightarrow \mathbb{R}^M$  defined by

$$\Upsilon(f) = \{P_j f(x_0, t_j)\}_{j=1, \dots, J} \in \mathbb{R}^M, \quad M = \sum_{j=1}^J q_j. \quad (10)$$

This functional corresponds to the discrete partial observations  $P_j X(t_j)$  of the diffusion process  $X(t) = \phi_{x_0}^{0,t}(B)$ . To invoke the calculus of variations we state the estimation problem in terms of  $\widehat{X}(t) = \phi_{x_0}^{0,t}(\widehat{B})$  and  $\widehat{X}^{(\epsilon)}(t) = \phi_{x_0}^{0,t}(\widehat{B}^{(\epsilon)})$ .

**Lemma 1.** Assume condition (A2). If  $\widehat{B}$  exists, then

$$\partial_t \widehat{B}(t) = \sigma(u)^\top (\sigma(u)\sigma(u)^\top)^{-1} (\partial_t \widehat{X}(t) - b(u)) \big|_{u=\widehat{X}(t)}. \quad (11)$$

If  $\widehat{B}^{(\epsilon)}$  exists, then the similar relation holds true between  $\widehat{B}^{(\epsilon)}$  and  $\widehat{X}^{(\epsilon)}$ .

**Proof.** Let  $f \in \mathcal{H}$  be given. The orthogonal projection

$$g(t) = \int_0^t \sigma(u)^\top (\sigma(u)\sigma(u)^\top)^{-1} \sigma(u) \big|_{u=\phi_{x_0}^{0,s}(f)} df(s)$$

satisfies  $\|g\|_{\mathcal{H}} \leq \|f\|_{\mathcal{H}}$  and  $\phi_{x_0}^{0,t}(g) = \phi_{x_0}^{0,t}(f)$ . Thus,

$$\partial_t \widehat{B}^{(\epsilon)}(t) = \sigma(u)^\top (\sigma(u)\sigma(u)^\top)^{-1} \sigma(u) \big|_{u=\phi_{x_0}^{0,t}(\widehat{B}^{(\epsilon)})} \partial_t \widehat{B}^{(\epsilon)}(t),$$

whereby the lemma follows.  $\square$

**Theorem 5.** Assume conditions (A1), (A2) and (A3), and let the Lagrangian  $F: \mathbb{R}^q \times \mathbb{R}^q \rightarrow \mathbb{R}$  be defined by

$$F(u, p) = \frac{1}{2} (p - b(u))^\top (\sigma(u)\sigma(u)^\top)^{-1} (p - b(u)).$$

If the functional  $\Upsilon$  is given by eq. (10), then the estimators  $\widehat{X}(t) = \phi_{x_0}^{0,t}(\widehat{B})$  and  $\widehat{X}^{(\epsilon)}(t) = \phi_{x_0}^{0,t}(\widehat{B}^{(\epsilon)})$  exist and are of class  $\mathcal{C}^2$  separately on the intervals  $(t_{j-1}, t_j)$ , where they satisfy the Euler-Lagrange equation

$$\frac{d}{dt} D_p F(\widehat{X}^{(\epsilon)}(t), \partial_t \widehat{X}^{(\epsilon)}(t)) = D_u F(\widehat{X}^{(\epsilon)}(t), \partial_t \widehat{X}^{(\epsilon)}(t)). \quad (12)$$

Using the convention  $\partial_t \widehat{B}(t+) = \partial_t \widehat{B}^{(\epsilon)}(t+) = 0$  for  $t = T$  we have

$$\begin{aligned} & \epsilon^{-2} P_j^\top (P_j \widehat{X}^{(\epsilon)}(t_j) - P_j X(t_j)) \\ &= (\sigma(z)\sigma(z)^\top)^{-1} \sigma(z) (\partial_t \widehat{B}^{(\epsilon)}(t+) - \partial_t \widehat{B}^{(\epsilon)}(t-)) \big|_{t=t_j, z=\widehat{X}^{(\epsilon)}(t_j)}. \end{aligned} \quad (13)$$

If the estimator  $\widehat{B}$  is uniquely defined, then  $\widehat{B}^{(\epsilon)} \rightarrow \widehat{B}$  as elements of  $\mathcal{H}$  as  $\epsilon \rightarrow 0$ . In particular, the limit of eq. (13) exists as  $\epsilon \rightarrow 0$ .

**Proof.** In this proof we write  $\widehat{B}^{(0)}, \widehat{X}^{(0)}$  for  $\widehat{B}, \widehat{X}$  when convenient. Let  $\mathcal{H}_X$  be the space of functions  $f \in \mathcal{C}([0, T]; \mathbb{R}^q)$  with  $f(0) = x_0$  and derivatives  $\partial f \in L^2([0, T]; \mathbb{R}^q)$ , and let the energy functional  $E_X: \mathcal{H}_X \rightarrow \mathbb{R}$  be defined by

$$E_X(f) = \int_0^T F(f(t), \partial_t f(t)) dt, \quad f \in \mathcal{H}_X.$$

If the estimators  $\widehat{B}^{(\epsilon)}$  exist and we define  $x_j^{(\epsilon)} = \phi_{x_0}^{0,t_j}(\widehat{B}^{(\epsilon)})$ , then Lemma 1 implies  $\|\widehat{B}^{(\epsilon)}\|_{\mathcal{H}}^2 = E_X(\widehat{X}^{(\epsilon)})$  and that  $\widehat{X}^{(\epsilon)}$  minimize  $E_X(f)$  over  $f \in \mathcal{H}_X$

with boundary conditions  $f(t_j) = x_j^{(\epsilon)}$ . Conversely, suppose there exists minimizers  $\widehat{X}^{(\epsilon)}$  for  $E_X(f)$  over  $f \in \mathcal{H}_X$  with boundary conditions  $f(t_j) = x_j^{(\epsilon)}$ . Choosing the points  $x_j^{(\epsilon)}$  via an optimization over the finite dimensional space  $\mathbb{R}^{q \times J}$  we may define the estimators  $\widehat{B}^{(\epsilon)}$  by eq. (11). Thus, the optimization problem for  $\widehat{B}^{(\epsilon)}$  can be recast in terms of  $\widehat{X}^{(\epsilon)}$ ,  $\epsilon \geq 0$ . To show the existence of the minimizers  $\widehat{X}^{(\epsilon)}$  we appeal to the so-called direct method in the calculus of variations. This technique was devised by Tonelli [22] and rely on the demonstration of lower semi-continuity of  $E_X$ . Let  $\widehat{X}_k^{(\epsilon)} \in \mathcal{H}_X$  be a sequence of approximate minimizers, i.e.  $P_j \widehat{X}_k^{(\epsilon)}(t_j) = P_j X(t_j)$  for  $j = 1, \dots, J$  and

$$\lim_{k \rightarrow \infty} E_X(\widehat{X}_k^{(\epsilon)}) = \inf_{f \in \mathcal{H}_X : P_j f(t_j) = P_j X(t_j)} E_X(f).$$

The assumptions (A1), (A2) and (A3) imply that the Lagrangian  $F(u, p)$  is autonomous, coersive, convex, and of class  $\mathcal{C}^2$  on the intervals  $(t_{j-1}, t_j)$ , and that  $D_p^2 F(u, p)$  is positive definite. Hence the Tonelli theorems [7, Theorem 1, Theorem 2] imply that the limit  $\widehat{X}^{(\epsilon)} = \lim_{k \rightarrow \infty} \widehat{X}_k^{(\epsilon)}$  exists and is a solution of the minimization problem. Moreover,  $\widehat{X}^{(\epsilon)}$  is of class  $\mathcal{C}^2$  and satisfies eq. (12) separately on the intervals  $(t_{j-1}, t_j)$ . This allows us to apply integration by parts on the second term in first variation  $\delta_f E_X(f)$  of the functional  $E_X$ ,

$$\delta_f E_X(f) \cdot g = \int_0^T D_u F(f(t), \partial_t f(t)) g(t) dt + \int_0^T D_p F(f(t), \partial_t f(t)) \partial_t g(t) dt.$$

Thus, for every  $g \in \mathcal{H}$  we have  $g(0) = 0$  and

$$\begin{aligned} 0 &= \delta_f \left( \frac{1}{2\epsilon^2} \sum_{j=1}^J |P_j f(t_j) - P_j X(t_j)|^2 + E_X(f) \right) \Big|_{f=\widehat{X}^{(\epsilon)}} \cdot g \\ &= \epsilon^{-2} \sum_{j=1}^J (P_j \widehat{X}^{(\epsilon)}(t_j) - P_j X(t_j))^\top P_j g(t_j) \\ &\quad + \sum_{j=1}^{J-1} (\partial_t \widehat{X}^{(\epsilon)}(t_j-) - \partial_t \widehat{X}^{(\epsilon)}(t_j+))^\top (\sigma(u)\sigma(u)^\top)^{-1} \Big|_{u=\widehat{X}^{(\epsilon)}(t_j)} g(t_j) \\ &\quad + (\partial_t \widehat{X}^{(\epsilon)}(T) - b(u))^\top (\sigma(u)\sigma(u)^\top)^{-1} \Big|_{u=\widehat{X}^{(\epsilon)}(T)} g(T), \end{aligned}$$

whereby eq. (13) follows. Finally, we consider the relation between  $\widehat{X}^{(\epsilon)}$  and  $\widehat{X}$ . Clearly, we have  $P_j \widehat{X}^{(\epsilon)}(t_j) \rightarrow P_j \widehat{X}(t_j) = P_j X(t_j)$  and  $E_X(\widehat{X}^{(\epsilon)}) \rightarrow E_X(\widehat{X})$  as  $\epsilon \rightarrow 0$ . Thus, if  $\widehat{X}(t) = \phi_{x_0}^{0,t}(\widehat{B})$  is uniquely defined, then the lower semi-continuity of  $E_X$  implies  $\widehat{X}^{(\epsilon)} \rightarrow \widehat{X}$  and  $\widehat{B}^{(\epsilon)} \rightarrow \widehat{B}$  as  $\epsilon \rightarrow 0$ .  $\square$

The statements of Theorem 4 and 5 assume uniqueness of the MAP estimators  $\widehat{B}^{(\epsilon)}$  and  $\widehat{B}$ . It is, however, easy to construct examples where these

estimators are non-unique. The prototypical examples of non-uniqueness involve set-ups with identical energies for going “left” or “right” around some center point. To remedy the defect arising from non-uniqueness of the estimators and retain the validity of the diagram (9) we could define  $\widehat{B}_N^{(\epsilon)}$  and  $\widehat{B}^{(\epsilon)}$  in a neighborhood of one particular  $\widehat{B}$ . But to keep the presentation as clear as possible we have omitted these technicalities.

## 4 Structure of the curvature operator

In Section 5 we find the Laplace approximation of the Lebesgue density for the random variable  $Y = (\Upsilon \circ \Phi)(B)$  as in eq. (10). The program is to integrate over  $\partial f$  in the equations (6), see eq. (18) below. To prepare this analysis we study the functional  $E: \mathcal{K} \rightarrow \mathbb{R}$  for fixed  $y = \{y_j\}_{j=1,\dots,J} \in \mathbb{R}^M$  defined by

$$E(\partial f) = -\epsilon^2 \log p^{(\epsilon)}(y|f) = \frac{1}{2} \sum_{j=1}^J |P_j \phi_{x_0}^{0,t_j}(f) - y_j|^2. \quad (14)$$

To describe the variational derivatives of  $E$  we invoke product integrals [10] and some tensor notation. For a function  $f = \{f_j\}_{j=1,\dots,q} \in \mathcal{C}^2(\mathbb{R}^q; \mathbb{R}^q)$  of the argument  $z = \{z_i\}_{i=1,\dots,q} \in \mathbb{R}^q$  the tensors  $\nabla_z f(z), \nabla_z^\top f(z) \in \mathbb{R}^{q \times q}$  and  $\nabla_z \nabla_z^\top f(z) \in \mathbb{R}^{q \times q \times q}$  are defined by

$$\nabla_z f(z) = \{\partial_{z_i} f_j(z)\}_{ji}, \quad \nabla_z^\top f(z) = \{\partial_{z_i} f_j(z)\}_{ij}, \quad \nabla_z \nabla_z^\top f(z) = \{\partial_{z_k} \partial_{z_i} f_j(z)\}_{ijk}.$$

Standard matrix notion is used for left and right multiplication. Furthermore, for  $x = \{x_j\}_{j=1,\dots,q} \in \mathbb{R}^{1 \times q}$  and  $z = \{z_{ijk}\}_{i,j,k=1,\dots,q} \in \mathbb{R}^{q \times q \times q}$  we write

$$x \star_2 z = \left\{ \sum_{j=1}^q x_j z_{ijk} \right\}_{i,k=1,\dots,q} \in \mathbb{R}^{q \times q}.$$

**Proposition 2.** *The second variation of  $E(\partial f)$  w.r.t.  $\partial f \in \mathcal{K}$  in the directions  $g, h \in \mathcal{K}$  is given by*

$$\begin{aligned} \delta_{\partial f}^2 E(\partial f) \cdot (g, h) &= \int_0^T \int_0^T g(t)^\top \beta_f(t) \alpha_f(t \vee s) \beta_f(s)^\top h(s) \, ds \, dt \\ &\quad + \int_0^T \int_0^T g(t)^\top \gamma_f(t) \gamma_f(s)^\top h(s) \, ds \, dt, \end{aligned}$$

where  $\alpha_f(t) \in \mathbb{R}^{q \times q}$ ,  $\beta_f(t) \in \mathbb{R}^{d \times q}$ ,  $\gamma_f(t) \in \mathbb{R}^{d \times M}$  for  $t \in [0, T)$  are defined

by

$$\begin{aligned}\alpha_f(t) &= \sum_{j:t_j \geq t} \left( (P_j \phi_{x_0}^{0,t_j}(f) - y_j)^\top P_j \kappa_f(t_j)^\top \right) \star_2 (\xi_f(t) + \lambda_f(t_j) - \lambda_f(t)), \\ \beta_f(t) &= \sigma(\phi_{x_0}^{0,t}(f))^\top \kappa_f(t)^{-1}, \\ \gamma_f(t) &= \left\{ 1_{t \leq t_j} \sigma(\phi_{x_0}^{0,t}(f))^\top \kappa_f(t)^{-1} \kappa_f(t_j) P_j^\top \right\}_{j=1, \dots, J}\end{aligned}$$

with  $\kappa_f(t) \in \mathbb{R}^{q \times q}$ ,  $\xi_f(t) \in \mathbb{R}^{q \times q \times q}$ ,  $\lambda_f(t) \in \mathbb{R}^{q \times q \times q}$  for  $t \in [0, T]$  defined by

$$\begin{aligned}\kappa_f(t) &= \prod_{s \in (0, t]} \left( \mathbb{I}_{\mathbb{R}^q} + \nabla_z^\top (b(z) + \sigma(z) \partial_s f(s)) \Big|_{z=\phi_{x_0}^{0,s}(f)} ds \right), \\ \xi_f(t) &= \kappa_f(t) \nabla_z \left( \kappa_f(t)^{-1, \top} \sigma(z) \sigma(w)^\top (\sigma(w) \sigma(w)^\top)^{-1} \right) \Big|_{z=w=\phi_{x_0}^{0,t}(f)} \kappa_f(t)^\top, \\ \lambda_f(t) &= \int_0^t \kappa_f(s) \nabla_z \nabla_z^\top \left( \kappa_f(s)^{-1, \top} (b(z) + \sigma(z) \partial_s f(s)) \right) \Big|_{z=\phi_{x_0}^{0,s}(f)} \kappa_f(s)^\top ds.\end{aligned}$$

Moreover, we define  $\alpha_f(T) = 0$ .

**Proof.** First we find representations for the first and second variation of  $\phi_{x_0}^{0,t}(f)$  w.r.t.  $\partial f \in \mathcal{K}$ . Concerning the first variation we have

$$\partial_t \phi_{x_0}^{0,t}(f) = b(\phi_{x_0}^{0,t}(f)) + \sigma(\phi_{x_0}^{0,t}(f)) \partial_t f(t),$$

and hence  $\partial_t \delta_{\partial f} \phi_{x_0}^{0,t}(f) \cdot g$  with  $g \in \mathcal{K}$  equals

$$\sigma(\phi_{x_0}^{0,t}(f)) g(t) + \nabla_z (b(z) + \sigma(z) \partial_t f(t)) \Big|_{z=\phi_{x_0}^{0,t}(f)} \delta_{\partial f} \phi_{x_0}^{0,t}(f) \cdot g.$$

This linear ODE is solved via the product integral [10, Theorem 10] by

$$\delta_{\partial f} \phi_{x_0}^{0,t}(f) \cdot g = \int_0^t \kappa_f(t)^\top \kappa_f(s)^{-1, \top} \sigma(\phi_{x_0}^{0,s}(f)) g(s) ds.$$

The second variation  $\delta_{\partial f}^2 \phi_{x_0}^{0,t}(f) \cdot (g, h)$  with  $g, h \in \mathcal{K}$  equals

$$\begin{aligned}& \int_0^t \kappa_f(t)^\top \kappa_f(s)^{-1, \top} \delta_{\partial f} \left( \sigma(\phi_{x_0}^{0,s}(f)) g(s) \right) \cdot h ds \\ & + \int_0^t \left( \delta_{\partial f} (\kappa_f(s)^{-1} \kappa_f(t)) \cdot h \right)^\top \sigma(\phi_{x_0}^{0,s}(f)) g(s) ds.\end{aligned} \tag{15}$$

The first term in eq. (15) equals

$$\int_0^t \int_0^s \kappa_f(t)^\top \nabla_z \left( \kappa_f(s)^{-1, \top} \sigma(z) g(s) \right) \Big|_{z=\phi_{x_0}^{0,s}(f)} \kappa_f(s)^\top \kappa_f(u)^{-1, \top} \sigma(\phi_{x_0}^{0,u}(f)) h(u) du ds. \tag{16}$$

The Duhamel equation [10, Theorem 4] gives that  $\delta_{\partial f}(\kappa_s(t)^{-1}\kappa_f(t)) \cdot h$  equals

$$\begin{aligned}
& \delta_{\partial f} \left\{ \prod_{u \in (s, t]} \left( \mathbb{I} + \nabla_z^\top (b(z) + \sigma(z) \partial_u f(u)) \Big|_{z=\phi_{x_0}^{0,u}(f)} du \right) \right\} \cdot h \\
&= \int_s^t \kappa_f(s)^{-1} \kappa_f(u) \left\{ \delta_{\partial f} \nabla_z^\top (b(z) + \sigma(z) \partial_u f(u)) \cdot h \right\} \kappa_f(u)^{-1} \kappa_f(t) \Big|_{z=\phi_{x_0}^{0,u}(f)} du \\
&= \int_s^t \kappa_f(s)^{-1} \kappa_f(u) \nabla_z^\top \left( \sigma(z) h(u) \right) \Big|_{z=\phi_{x_0}^{0,u}(f)} \kappa_f(u)^{-1} \kappa_f(t) du \\
&\quad + \int_s^t \int_0^u \kappa_f(s)^{-1} \kappa_f(u) \left( \nabla_z \nabla_z^\top (b(z) + \sigma(z) \partial_u f(u)) \Big|_{z=\phi_{x_0}^{0,u}(f)} \right. \\
&\quad \left. \kappa_f(u)^\top \kappa_f(v)^{-1, \top} \sigma(\phi_{x_0}^{0,v}(f)) h(v) \right) \kappa_f(u)^{-1} \kappa_f(t) dv du,
\end{aligned}$$

and hence the second term in eq. (15) equals

$$\begin{aligned}
& \int_0^t \int_s^t \kappa_f(t)^\top \nabla_z (\kappa_f(u)^{-1, \top} \sigma(z) h(u)) \Big|_{z=\phi_{x_0}^{0,u}(f)} \kappa_f(u)^\top \kappa_f(s)^{-1, \top} \sigma(\phi_{x_0}^{0,s}(f)) g(s) du ds \\
&\quad + \kappa_f(t)^\top \int_0^t \int_s^t \int_0^u h(v)^\top \sigma(\phi_{x_0}^{0,v}(f))^\top \kappa_f(v)^{-1} \kappa_f(u) \nabla_z \nabla_z^\top \left( \kappa_f(u)^{-1, \top} \right. \\
&\quad \left. (b(z) + \sigma(z) \partial_u f(u)) \right) \Big|_{z=\phi_{x_0}^{0,u}(f)} \kappa_f(u)^\top \kappa_f(s)^{-1, \top} \sigma(\phi_{x_0}^{0,s}(f)) g(s) dv du ds.
\end{aligned} \tag{17}$$

If the orthogonal projection  $\pi_f: \mathcal{K} \rightarrow \mathcal{K}$  is defined by

$$(\pi_f g)(t) = \sigma(z)^\top (\sigma(z) \sigma(z)^\top)^{-1} \sigma(z) \Big|_{z=\phi_{x_0}^{0,t}(f)} g(t),$$

then  $\delta_{\partial f}^2 \phi_{x_0}^{0,t}(f) \cdot (g, h) = \delta_{\partial f}^2 \phi_{x_0}^{0,t}(f) \cdot (\pi_f g, \pi_f h)$ . Employing this identity and combining eq. (16) and eq. (17) we have that  $\delta_{\partial f}^2 \phi_{x_0}^{0,t}(f) \cdot (g, h)$  equals

$$\begin{aligned}
& \kappa_f(t)^\top \int_0^t \int_0^s g(s)^\top \sigma(\phi_{x_0}^{0,s}(f))^\top \kappa_f(s)^{-1} \left( \xi_f(s \vee u) + \lambda_f(t) - \lambda_f(s \vee u) \right) \\
& \quad \kappa_f(u)^{-1, \top} \sigma(\phi_{x_0}^{0,s}(f)) h(u) du ds.
\end{aligned}$$

Since the second variation  $\delta_{\partial f}^2 E(\partial f) \cdot (g, h)$  equals

$$\begin{aligned}
& \sum_{j=1}^J (P_j \phi_{x_0}^{0,t_j}(f) - y_j)^\top P_j \delta_{\partial f}^2 \phi_{x_0}^{0,t_j}(f) \cdot (g, h) \\
& + \sum_{j=1}^J (P_j \delta_{\partial f} \phi_{x_0}^{0,t_j}(f) \cdot g)^\top (P_j \delta_{\partial f} \phi_{x_0}^{0,t_j}(f) \cdot h),
\end{aligned}$$

the proposition follows inserting the above formulae.  $\square$

## 5 Path integrals

In this section we find the Laplace approximation of the Lebesgue density of the discrete partial observation described via eq. (10). The random variables

$$Y_N^{(\epsilon)} = (\Upsilon \circ \Phi)(B_N) + \epsilon R, \quad Y^{(\epsilon)} = (\Upsilon \circ \Phi)(B) + \epsilon R$$

satisfy  $\lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} Y_N^{(\epsilon)} = (\Upsilon \circ \Phi)(B)$  in probability. Thus, the limit of the Laplace approximation of the Lebesgue density  $p_N^{(\epsilon)}(y)$  for  $Y_N^{(\epsilon)} = y \in \mathbb{R}^M$  provides an approximation of the Lebesgue density for  $(\Upsilon \circ \Phi)(B) = y$ . In the limit as  $N \rightarrow \infty$  we essentially compute the integral over all Brownian paths. Since the Brownian motion has independent increments this limit is more naturally described as an integral over  $\mathcal{K}$  than over  $\mathcal{H}$ . To approximate the space  $\mathcal{K}$  by  $\mathbb{R}^{d \times N}$  we introduce the operators  $W_N: \mathbb{R}^{d \times N} \rightarrow \mathcal{K}$  defined by

$$W_N(u)(t) = \frac{N}{T} u_{\lceil \frac{tN}{T} \rceil}, \quad u = \{u_n\}_{n=1, \dots, N} \in \mathbb{R}^{d \times N}, \quad N \in \mathbb{N}.$$

Observe that  $\frac{T}{N} W_N^\top W_N$  is the identity operator on  $\mathbb{R}^{d \times N}$  and that  $\frac{T}{N} W_N W_N^\top$  is the projection of  $\mathcal{K}$  on the subspace  $W_N(\mathbb{R}^{d \times N})$  of functions that are piecewise constant on the intervals  $(\frac{(n-1)T}{N}, \frac{nT}{N})$ . Invoking the functional  $E: \mathcal{K} \rightarrow \mathbb{R}$  from eq. (14) and the density of  $\partial B_N$  w.r.t.  $W_N(\text{Leb}(\mathbb{R}^{d \times N}))$  we have that the Lebesgue density for  $Y_N^{(\epsilon)} = y$  is given by

$$p_N^{(\epsilon)}(y) = (2\pi\epsilon^2)^{-M/2} \left( \frac{N}{2\pi T} \right)^{-Nd/2} \int_{\mathbb{R}^{d \times N}} e^{-\frac{1}{2} \|W_N(u)\|_{\mathcal{K}}^2 - \epsilon^{-2} E(W_N(u))} du. \quad (18)$$

The Laplace approximation of this integral is defined via the second order Taylor approximation of the negative exponent around its minimizer. Proposition 2 provides the second variation  $\epsilon^{-2} \delta_{\partial f}^2 E(\partial \hat{B}_N^{(\epsilon)}) = \hat{A}_N^{(\epsilon)} + \hat{C}_N^{(\epsilon)}$  with

$$\begin{aligned} \hat{A}_N^{(\epsilon)} g(t) &= \epsilon^{-2} \int_0^T \beta_f(t) \alpha_f(t \vee s) \beta_f(s)^\top g(s) ds \Big|_{f=\hat{B}_N^{(\epsilon)}}, \\ \hat{C}_N^{(\epsilon)} g(t) &= \epsilon^{-2} \int_0^T \gamma_f(t) \gamma_f(s)^\top g(s) ds \Big|_{f=\hat{B}_N^{(\epsilon)}}. \end{aligned}$$

The estimator  $\partial \hat{B}_N^{(\epsilon)} \in \mathcal{K}$  is an approximate minimizer for  $\frac{1}{2} \|\partial f\|_{\mathcal{K}}^2 + \epsilon^{-2} E(\partial f)$  over  $\partial f \in W_N(\mathbb{R}^{d \times N})$ . Thus,  $\frac{T}{N} W_N^\top (\partial \hat{B}_N^{(\epsilon)}) \in \mathbb{R}^{d \times N}$  is an approximate minimizer for the negative exponent of the integrand in eq. (18). The associated second order Taylor approximation is given by

$$\begin{aligned} \frac{1}{2} \|W_N(u)\|_{\mathcal{K}}^2 + \epsilon^{-2} E(W_N(u)) &\approx \frac{1}{2} \|\partial \hat{B}_N^{(\epsilon)}\|_{\mathcal{K}}^2 + \epsilon^{-2} E(\partial \hat{B}_N^{(\epsilon)}) \\ &+ \frac{1}{2} (u - \hat{u})^\top \left( W_N^\top W_N + W_N^\top (\hat{A}_N^{(\epsilon)} + \hat{C}_N^{(\epsilon)}) W_N \right) (u - \hat{u}) \Big|_{\hat{u} = \frac{T}{N} W_N^\top (\partial \hat{B}_N^{(\epsilon)})}, \end{aligned}$$



and the associated Laplace approximation of the integral eq. (18) is given by

$$\begin{aligned}\hat{p}_N^{(\epsilon)}(y) &= (2\pi\epsilon^2)^{-M/2} \left(\frac{N}{2\pi T}\right)^{-Nd/2} \exp\left(-\frac{\|\partial\hat{B}_N^{(\epsilon)}\|_{\mathcal{K}}^2}{2} - \epsilon^{-2}E(\partial\hat{B}_N^{(\epsilon)})\right) \\ &\quad \int_{\mathbb{R}^{d \times N}} \exp\left(-\frac{u^\top (\mathbf{W}_N^\top \mathbf{W}_N + \mathbf{W}_N^\top (\hat{A}_N^{(\epsilon)} + \hat{C}_N^{(\epsilon)}) \mathbf{W}_N) u}{2}\right) du \\ &= (2\pi)^{-M/2} \exp\left(-\frac{\|\partial\hat{B}_N^{(\epsilon)}\|_{\mathcal{K}}^2}{2} - \epsilon^{-2}E(\partial\hat{B}_N^{(\epsilon)})\right) \\ &\quad \left\{ \epsilon^{2M} \det\left(\mathbb{I}_{\mathbb{R}^{d \times N}} + \frac{T}{N} \mathbf{W}_N^\top (\hat{A}_N^{(\epsilon)} + \hat{C}_N^{(\epsilon)}) \mathbf{W}_N\right) \right\}^{-1/2}.\end{aligned}$$

Let  $\hat{A}^{(\epsilon)}$  and  $\hat{C}^{(\epsilon)}$  be defined similarly as  $\hat{A}_N^{(\epsilon)}$  and  $\hat{C}_N^{(\epsilon)}$ . Eq. (13) implies that  $\lim_{\epsilon \rightarrow 0} \epsilon^{-2} \alpha_{\hat{B}^{(\epsilon)}}(t)$  and  $\hat{A} = \lim_{\epsilon \rightarrow 0} \hat{A}^{(\epsilon)}$  exist. Let  $\hat{Q}$  be the projection of  $\mathcal{K}$  on the orthogonal complement of the subspace spanned by  $\gamma_f$  with  $f = \hat{B}$ , i.e.

$$\hat{Q}g(t) = g(t) - \gamma_f(t) \left( \int_0^T \gamma_f(s)^\top \gamma_f(s) ds \right)^{-1} \int_0^T \gamma_f(s)^\top g(s) ds \Big|_{f=\hat{B}}, \quad g \in \mathcal{K},$$

and let  $\hat{Q}^{(\epsilon)}$  be defined similarly using  $\gamma_f$  with  $f = \hat{B}^{(\epsilon)}$ .

**Theorem 6.** *Assume conditions (A1) and (A2). If  $\hat{B}, \hat{B}^{(\epsilon)} \in \mathcal{H}$  exist and  $\hat{B}^{(\epsilon)} \rightarrow \hat{B}$  as  $\epsilon \rightarrow 0$ , then  $\lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} (-2 \log \hat{p}_N^{(\epsilon)}(y))$  exists and equals*

$$\begin{aligned} & M \log(2\pi) + \|\partial\hat{B}\|_{\mathcal{K}}^2 \\ & + \log \det \left( \int_0^T \gamma_f(s)^\top \gamma_f(s) ds \Big|_{f=\hat{B}} \right) + \text{tr}_{\mathcal{K}} [\log (\mathbb{I}_{\mathcal{K}} + \hat{Q}\hat{A}\hat{Q})]. \end{aligned} \quad (19)$$

Furthermore, if the operators  $\hat{G}: \mathcal{K} \rightarrow \mathbb{R}^M$  and  $\hat{G}^\top: \mathbb{R}^M \rightarrow \mathcal{K}$  are defined by

$$\hat{G}g = \int_0^T \gamma_f(s)^\top g(s) ds \Big|_{f=\hat{B}}, \quad (\hat{G}^\top u)(t) = \gamma_f(t)u \Big|_{f=\hat{B}},$$

then  $\text{tr}_{\mathcal{K}}[\log(\mathbb{I}_{\mathcal{K}} + \hat{Q}\hat{A}\hat{Q})]$  equals

$$\int_0^1 \left( \text{tr}_{\mathcal{K}}[\hat{H}_z \hat{A}] - \text{tr}_{\mathbb{R}^M}[(\hat{G} \hat{H}_z \hat{G}^\top)^{-1} \hat{G} \hat{H}_z \hat{A} \hat{H}_z \hat{G}^\top] \right) \Big|_{\hat{H}_z = (\mathbb{I}_{\mathcal{K}} + z\hat{A})^{-1}} dz. \quad (20)$$

**Proof.** We have that  $-2 \log \hat{p}_N^{(\epsilon)}(y)$  equals

$$\begin{aligned} & M \log(2\pi) + \|\partial\hat{B}_N^{(\epsilon)}\|_{\mathcal{K}}^2 + 2\epsilon^{-2}E(\partial\hat{B}_N^{(\epsilon)}) \\ & + \log \left( \epsilon^{2M} \det \left( \mathbb{I}_{\mathbb{R}^{d \times N}} + \frac{T}{N} \mathbf{W}_N^\top (\hat{A}_N^{(\epsilon)} + \hat{C}_N^{(\epsilon)}) \mathbf{W}_N \right) \right).\end{aligned}$$

Theorem 4, Theorem 5 and eq. (13) imply

$$\lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} \|\partial \widehat{B}_N^{(\epsilon)}\|_{\mathcal{K}}^2 = \|\partial \widehat{B}\|_{\mathcal{K}}^2, \quad \lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} \epsilon^{-2} E(\partial \widehat{B}_N^{(\epsilon)}) = 0.$$

Regularity condition (A2) implies that the matrix

$$\frac{T}{N} \mathbf{W}_N^\top \widehat{C}_N^{(\epsilon)} \mathbf{W}_N = \epsilon^{-2} \frac{T}{N} (\mathbf{W}_N^\top \gamma_f) (\mathbf{W}_N^\top \gamma_f)^\top \Big|_{f=\widehat{B}_N^{(\epsilon)}} \in \mathbb{R}^{(d \times N) \times (d \times N)}$$

has rank  $M$  for  $N$  sufficiently large. If the operator  $\widehat{U}_N^{(\epsilon)}$  on  $\mathcal{K}$  is defined by

$$\widehat{U}_N^{(\epsilon)} g(t) = \gamma_f(t) \left( \frac{T}{N} (\mathbf{W}_N^\top \gamma_f) (\mathbf{W}_N^\top \gamma_f)^\top \right)^{-3/2} \int_0^T \gamma_f(s)^\top g(s) ds \Big|_{f=\widehat{B}_N^{(\epsilon)}}$$

and  $\widehat{Q}_N^{(\epsilon)}$  is the projection of  $\mathcal{K}$  on the orthogonal complement of the  $M$ -dimensional subspace spanned by  $\frac{T}{N} \mathbf{W}_N \mathbf{W}_N^\top \gamma_f$  for  $f = \widehat{B}_N^{(\epsilon)}$ , then

$$\left( \frac{T}{N} \mathbf{W}_N^\top (\widehat{C}_N^{(\epsilon)} + \widehat{Q}_N^{(\epsilon)}) \mathbf{W}_N \right)^{-1/2} = \frac{T}{N} \mathbf{W}_N^\top (\epsilon \widehat{U}_N^{(\epsilon)} + \widehat{Q}_N^{(\epsilon)}) \mathbf{W}_N.$$

Thus, since  $\frac{T}{N} \mathbf{W}_N^\top \widehat{C}_N^{(\epsilon)} \mathbf{W}_N$  has rank  $M$  we have that

$$\epsilon^{2M} \det \left( \mathbb{I}_{\mathbb{R}^{d \times N}} + \frac{T}{N} \mathbf{W}_N^\top (\widehat{A}_N^{(\epsilon)} + \widehat{C}_N^{(\epsilon)}) \mathbf{W}_N \right)$$

equals the determinant of

$$\begin{aligned} & \left( \frac{T}{N} \mathbf{W}_N^\top (\epsilon^2 \widehat{C}_N^{(\epsilon)} + \widehat{Q}_N^{(\epsilon)}) \mathbf{W}_N \right) \left( \frac{T}{N} \mathbf{W}_N^\top (\epsilon \widehat{U}_N^{(\epsilon)} + \widehat{Q}_N^{(\epsilon)}) \mathbf{W}_N \right) \\ & \left( \mathbb{I}_{\mathbb{R}^{d \times N}} + \frac{T}{N} \mathbf{W}_N^\top (\widehat{A}_N^{(\epsilon)} + \widehat{C}_N^{(\epsilon)}) \mathbf{W}_N \right) \left( \frac{T}{N} \mathbf{W}_N^\top (\epsilon \widehat{U}_N^{(\epsilon)} + \widehat{Q}_N^{(\epsilon)}) \mathbf{W}_N \right). \end{aligned} \quad (21)$$

For positive definite matrices  $\Xi \in \mathbb{R}^{(d \times N) \times (d \times N)}$  we have the identity

$$\begin{aligned} \det \Xi &= (\exp \circ \text{tr}_{\mathbb{R}^{d \times N}} \circ \log)(\Xi) \\ &= (\exp \circ \text{tr}_{\mathcal{K}} \circ \log) \left( \mathbb{I}_{\mathcal{K}} + \frac{T}{N} \mathbf{W}_N (\Xi - \mathbb{I}_{\mathbb{R}^{d \times N}}) \mathbf{W}_N^\top \right), \end{aligned}$$

where the operator logarithm is defined via spectral calculus [18, Chapter 4]. This formula allows us to compute the determinant of eq. (21) over the same space for all  $N \in \mathbb{N}$ , namely  $\mathcal{K}$ . Operator multiplication is sequentially continuous [18, Section 4.6.1] and  $\frac{T}{N} \mathbf{W}_N \mathbf{W}_N^\top \rightarrow \mathbb{I}_{\mathcal{K}}$ ,  $\widehat{C}_N^{(\epsilon)} \rightarrow \widehat{C}^{(\epsilon)}$ ,  $\widehat{Q}_N^{(\epsilon)} \rightarrow \widehat{Q}^{(\epsilon)}$  as  $N \rightarrow \infty$ . Thus, the determinant of the first factor in eq. (21) equals

$$\begin{aligned} & \exp \left( \text{tr}_{\mathcal{K}} \left[ \log \left( \mathbb{I}_{\mathcal{K}} + \frac{T^2}{N^2} \mathbf{W}_N \mathbf{W}_N^\top (\epsilon^2 \widehat{C}_N^{(\epsilon)} + \widehat{Q}_N^{(\epsilon)} - \mathbb{I}_{\mathbb{R}^{d \times N}}) \mathbf{W}_N \mathbf{W}_N^\top \right) \right] \right) \\ & \xrightarrow{N \rightarrow \infty} \exp \left( \text{tr}_{\mathcal{K}} \left[ \log \left( \epsilon^2 \widehat{C}^{(\epsilon)} + \widehat{Q}^{(\epsilon)} \right) \right] \right) = \det \left( \int_0^T \gamma_f(t)^\top \gamma_f(t) dt \Big|_{f=\widehat{B}^{(\epsilon)}} \right) \\ & \xrightarrow{\epsilon \rightarrow 0} \det \left( \int_0^T \gamma_f(t)^\top \gamma_f(t) dt \Big|_{f=\widehat{B}} \right). \end{aligned}$$

The product of the three last factors in eq. (21) equals

$$\begin{aligned} & \mathbb{I}_{\mathbb{R}^{d \times N}} + \epsilon^2 \frac{T^2}{N^2} \mathbf{W}_N^\top \widehat{U}_N^{(\epsilon)} \mathbf{W}_N \mathbf{W}_N^\top \widehat{U}_N^{(\epsilon)} \mathbf{W}_N \\ & + \epsilon \frac{T^2}{N^2} \mathbf{W}_N^\top \widehat{U}_N^{(\epsilon)} \mathbf{W}_N \mathbf{W}_N^\top \widehat{A}_N^{(\epsilon)} \mathbf{W}_N + \epsilon \frac{T^2}{N^2} \mathbf{W}_N^\top \widehat{A}_N^{(\epsilon)} \mathbf{W}_N \mathbf{W}_N^\top \widehat{U}_N^{(\epsilon)} \mathbf{W}_N \\ & + \frac{T^3}{N^3} \mathbf{W}_N^\top \widehat{Q}_N^{(\epsilon)} \mathbf{W}_N \mathbf{W}_N^\top \widehat{A}_N^{(\epsilon)} \mathbf{W}_N \mathbf{W}_N^\top \widehat{Q}_N^{(\epsilon)} \mathbf{W}_N. \end{aligned}$$

Invoking the operators  $\widehat{U}^{(\epsilon)} = \lim_{N \rightarrow \infty} \widehat{U}_N^{(\epsilon)}$  and continuity of the logarithm [18, Exercise 4.6.5] we have that the logarithm of the determinant of the three last factors in eq. (21) equals

$$\begin{aligned} & \text{tr}_{\mathcal{K}} \left[ \log \left( \mathbb{I}_{\mathcal{K}} + \epsilon^2 \frac{T^3}{N^3} \mathbf{W}_N \mathbf{W}_N^\top \widehat{U}_N^{(\epsilon)} \mathbf{W}_N \mathbf{W}_N^\top \widehat{U}_N^{(\epsilon)} \mathbf{W}_N \mathbf{W}_N^\top \right. \right. \\ & \quad + \epsilon \frac{T^3}{N^3} \mathbf{W}_N \mathbf{W}_N^\top \widehat{U}_N^{(\epsilon)} \mathbf{W}_N \mathbf{W}_N^\top \widehat{A}_N^{(\epsilon)} \mathbf{W}_N \mathbf{W}_N^\top \\ & \quad + \epsilon \frac{T^3}{N^3} \mathbf{W}_N \mathbf{W}_N^\top \widehat{A}_N^{(\epsilon)} \mathbf{W}_N \mathbf{W}_N^\top \widehat{U}_N^{(\epsilon)} \mathbf{W}_N \mathbf{W}_N^\top \\ & \quad \left. \left. + \frac{T^4}{N^4} \mathbf{W}_N \mathbf{W}_N^\top \widehat{Q}_N^{(\epsilon)} \mathbf{W}_N \mathbf{W}_N^\top \widehat{A}_N^{(\epsilon)} \mathbf{W}_N \mathbf{W}_N^\top \widehat{Q}_N^{(\epsilon)} \mathbf{W}_N \mathbf{W}_N^\top \right) \right] \\ & \xrightarrow{N \rightarrow \infty} \text{tr}_{\mathcal{K}} \left[ \log \left( \mathbb{I}_{\mathcal{K}} + \epsilon^2 \widehat{U}^{(\epsilon)} \widehat{U}^{(\epsilon)} + \epsilon \widehat{U}^{(\epsilon)} \widehat{A}^{(\epsilon)} + \epsilon \widehat{A}^{(\epsilon)} \widehat{U}^{(\epsilon)} + \widehat{Q}^{(\epsilon)} \widehat{A}^{(\epsilon)} \widehat{Q}^{(\epsilon)} \right) \right] \\ & \xrightarrow{\epsilon \rightarrow 0} \text{tr}_{\mathcal{K}} \left[ \log \left( \mathbb{I}_{\mathcal{K}} + \widehat{Q} \widehat{A} \widehat{Q} \right) \right]. \end{aligned}$$

This completes the proof of eq. (19). To prove eq. (20) we use

$$\begin{aligned} \text{tr}_{\mathcal{K}} \left[ \log \left( \mathbb{I}_{\mathcal{K}} + \widehat{Q} \widehat{A} \widehat{Q} \right) \right] &= \int_0^1 \text{tr}_{\mathcal{K}} \left[ \partial_z \log \left( \mathbb{I}_{\mathcal{K}} + z \widehat{Q} \widehat{A} \widehat{Q} \right) \right] dz \\ &= \int_0^1 \text{tr}_{\mathcal{K}} \left[ \left( \mathbb{I}_{\mathcal{K}} + z \widehat{Q} \widehat{A} \widehat{Q} \right)^{-1} \widehat{Q} \widehat{A} \widehat{Q} \right] dz. \end{aligned}$$

We have  $\mathbb{I}_{\mathcal{K}} - \widehat{Q} = \widehat{G}^\top (\int_0^T \gamma_f(s)^\top \gamma_f(s) ds)^{-1} \widehat{G}$  with  $f = \widehat{B}$ , and elementary matrix algebra gives that the trace of  $(\mathbb{I}_{\mathcal{K}} + z \widehat{Q} \widehat{A} \widehat{Q})^{-1} \widehat{Q} \widehat{A} \widehat{Q}$  equals

$$\left( \text{tr}_{\mathcal{K}} \left[ \widehat{H}_z \widehat{A} \right] - \text{tr}_{\mathbb{R}^M} \left[ (\widehat{G} \widehat{H}_z \widehat{G}^\top)^{-1} \widehat{G} \widehat{H}_z \widehat{A} \widehat{H}_z \widehat{G}^\top \right] \right) \Big|_{\widehat{H}_z = (\mathbb{I}_{\mathcal{K}} + z \widehat{A})^{-1}},$$

whereby eq. (20) follows.  $\square$

Theorem 7 stated below solves the Fredholm equation of the second kind  $(\mathbb{I}_{\mathcal{K}} + A)f = g$  for a class of operators  $A$  on  $\mathcal{K}$  given on the form

$$Af(t) = \int_0^T \beta(t) \alpha(t \vee s) \beta(s)^\top f(s) ds, \quad f \in \mathcal{K}. \quad (22)$$

Necessary and sufficient conditions for  $\mathbb{I}_{\mathcal{K}} + A$  to be invertible are given, and an accompanying inversion formula shows that the inverse operator also is given on the form eq. (22). The inversion formula can be invoked for the computation of the trace term eq. (20), where the operators  $\mathbb{I}_{\mathcal{K}} + z \widehat{A}$  are to be inverted. Proposition 2 implies that  $z \widehat{A}$  is given on the form

eq. (22) with  $\alpha(t) \in \mathbb{R}^{q \times q}$  and  $\beta(t) \in \mathbb{R}^{d \times q}$ . Combining these elements the Laplace approximation can be computed integrating systems of differential equations. In the appendix we explicitly state the systems to be solved in the univariate case  $q = d = 1$ .

To conclude this section we state the specialization of the results from [16] to symmetric operators.

**Theorem 7.** *Suppose that the operator  $A$  is given on the form eq. (22) with cadlag functions  $\alpha(t) \in \mathbb{R}^{K \times K}$  and  $\beta(t) \in \mathbb{R}^{d \times K}$  for some  $K \in \mathbb{N}$ , and where  $\alpha(t)$  also is continuous differentiable from the right with  $\alpha(T) = 0$ . For each  $t \in [0, T]$  let  $h_t(u) \in \mathbb{R}^{K \times K}$  be the unique solution to the Volterra integral equation of the second kind given by*

$$h_t(u) = 1_{u < t} \beta(u)^\top \beta(u) \left( \alpha(t) - \alpha(u) + \int_u^t (\alpha(s) - \alpha(u)) h_t(s) ds \right).$$

*Then  $\mathbb{I}_K + A$  is invertible if and only if the matrix  $\mathbb{I}_{\mathbb{R}^K} - \int_0^T h_T(s) ds \in \mathbb{R}^{K \times K}$  is invertible. Furthermore, if the matrices  $\mathbb{I}_{\mathbb{R}^K} - \int_0^t h_t(s) ds$  are invertible for  $t \in [0, T]$ , then  $(\mathbb{I}_K + A)^{-1} = \mathbb{I}_K + B$  with*

$$Bf(t) = \int_0^T \beta(t) \psi(t)^{-1} \rho(t \vee s) \chi(s)^{-1, \top} \beta(s)^\top f(s) ds.$$

*Here the cadlag functions  $\psi(t), \chi(t), \rho(t) \in \mathbb{R}^{K \times K}$  are given by*

$$\psi(t) = \prod_{s \in (0, t]} \left( \mathbb{I}_{\mathbb{R}^K} + (d\alpha(s)) \varphi(s) \right), \quad \chi(t) = \prod_{s \in (0, t]} \left( \mathbb{I}_{\mathbb{R}^K} + (d\alpha(s)^\top) \varphi(s)^\top \right)$$

*and  $\rho(t) = \int_{(t, T]} \psi(s-) (d\alpha(s)) \chi(s)^\top$ . The cadlag function  $\varphi(t) \in \mathbb{R}^{K \times K}$  solves the Riccati type jump-differential equation given by  $\varphi(0) = 0$  and*

$$\varphi(t) = \left( \mathbb{I}_{\mathbb{R}^K} - \varphi(t-) \Delta_t \alpha(t) \right)^{-1} \varphi(t-), \quad \partial_t \varphi(t) = \beta(t)^\top \beta(t) + \varphi(t) \partial_t \alpha(t) \varphi(t),$$

*where  $\Delta_t \alpha(t) = \alpha(t) - \alpha(t-)$  and  $\partial_t \alpha(t) = \lim_{\eta \rightarrow 0} \frac{\alpha(t+\eta) - \alpha(t)}{\eta}$  denote the jumps and right hand side derivatives of the function  $\alpha$ .*

## 6 Discussion

In this section we discuss various aspects of the developed theory. Before proceeding to the statistical aspects in the following subsections we comment on the relation to classical white noise theory. Here white noise is a stochastic Schwartz distribution defined as the generalized derivative of Brownian motion [12]. Streit & Hida [21] applied this theory for the analysis of the Feynman path integral and also suggested the Laplace approximation. Further investigations in this direction were made by Falco & Khandekar [9].

Albeit mathematically elegant, classical white noise theory is non-amenable for numerical computations. The advantage of our analysis in the Cameron-Martin space is that standard techniques from calculus of variations and for ordinary differential equations can be invoked. In particular, the Laplace approximation and the inversion formula stated in Theorem 6 and Theorem 7 are amenable for numerical computations.

### 6.1 Invariance of MAP paths and the Laplace approximation

We have studied MAP paths and the Laplace approximation of the density of a finite dimensional functional  $Y = (\Upsilon \circ \Phi)(B)$  of the Brownian motion. The MAP paths are invariant under non-linear transformations in the sense that the estimators do not depend on whether the transformation is used before or after the maximization. This invariance property is intuitively clear since the underlying MAP Brownian path remains unchanged. The Laplace approximation satisfies a similar invariance property: The Lebesgue density of  $Y$  is given by

$$p_Y(y) = \lim_{\epsilon \rightarrow 0} \mathbb{E} \left[ \frac{1}{(2\pi\epsilon^2)^{M/2}} \exp \left( - \frac{|Y - y|^2}{2\epsilon^2} \right) \right].$$

If  $\Psi: \mathbb{R}^M \rightarrow \mathbb{R}^M$  is a smooth diffeomorphism of the range space for  $Y$ , then the Lebesgue density of the transformed variable  $Z = \Psi(Y)$  is given by

$$\begin{aligned} p_Z(\Psi(y)) &= \lim_{\epsilon \rightarrow 0} \mathbb{E} \left[ \frac{1}{(2\pi\epsilon^2)^{M/2}} \exp \left( - \frac{|\Psi(Y) - \Psi(y)|^2}{2\epsilon^2} \right) \right] \\ &= \lim_{\epsilon \rightarrow 0} \mathbb{E} \left[ \frac{1}{(2\pi\epsilon^2)^{M/2}} \exp \left( - \frac{|\nabla \Psi(y) \cdot (Y - y)|^2}{2\epsilon^2} \right) \right] \\ &= \det(\nabla \Psi(y)^\top \nabla \Psi(y))^{-1/2} p_Y(y). \end{aligned}$$

This restatement of the transformation rule for Lebesgue densities remains true when the expectations over the Brownian motion are replaced by their Laplace approximations. Thus, if  $\hat{p}_Y(y)$  and  $\hat{p}_Z(z)$  denote the Laplace approximation of the densities for  $Y$  and  $Z$ , respectively, then

$$\hat{p}_Y(y) = \sqrt{\det(\nabla \Psi(y)^\top \nabla \Psi(y))} \hat{p}_Z(\Psi(y)). \quad (23)$$

As illustrated in Example 1 below the system of differential equations to be solved for the computation of  $\hat{p}_Y(y)$  and  $\hat{p}_Z(z)$  may be very different.

### 6.2 Comparison with conditioned stochastic differential equations

The MAP path of a stochastic differential equation  $X$  given its endpoints over the interval  $[0, T]$  can be compared to the conditional diffusion bridge

investigated by Baudoin [3]. Consider the univariate case  $q = d = 1$  and let the function  $\psi(t, x)$  be defined by the equation

$$\partial_t \widehat{X}(t) = b(\widehat{X}(t)) + \sigma(\widehat{X}(t))^2 \psi(t, \widehat{X}(t)).$$

Then  $\psi(t, x)$  satisfies the PDE  $\partial_t \psi + \partial_x(b \psi) + \frac{1}{2} \partial_x(\sigma^2 \psi^2) = 0$ . It is known [3, Proposition 37] that the diffusion bridge  $Z$  for  $X$  given the conditioning  $X(T) = x_T$  satisfies the conditioned SDE

$$dZ(t) = \left( \xi(Z(t)) + \sigma(Z(t))^2 \psi_0(t, Z(t)) \right) dt + \sigma(Z(t)) dW(t),$$

where  $\xi(x) = b(x) + \frac{1}{2} \sigma(x) \partial_x \sigma(x)$  is the Itô drift,  $W$  is a standard Brownian motion in the filtration enlarged by  $X(T)$ , and  $\psi_0(t, x)$  satisfies the PDE

$$\partial_t \psi_0 + \partial_x(\xi \psi_0) + \frac{1}{2} \partial_x(\sigma^2 \psi_0^2) + \frac{1}{2} \partial_x(\sigma^2 \partial_x \psi_0) = 0.$$

The PDEs for  $\psi(t, x)$  and  $\psi_0(t, x)$  exhibit similarities and can coincide for some instances of the coefficients  $b$  and  $\sigma$ . In particular, for linear functionals of the Brownian motion the MAP path coincides with the drift coefficient of the conditional diffusion bridge. Nevertheless, in general  $\psi(t, x)$  and  $\psi_0(t, x)$  differ in accordance with the distinctiveness of the maximum mode and the mean. The invariance property for the MAP path under non-linear transformations is not shared by the conditional diffusion bridge.

### 6.3 MAP and ML estimation

The developed theory can be applied for both MAP estimation of diffusion paths and for approximate maximum likelihood estimation of the parameter  $\theta \in \Theta$ . Below we present three simple examples in the univariate case  $q = d = M = 1$ . For univariate SDEs there exists other alternatives to exact MLE as reviewed by Sørensen [20]. Beskos et al. [5] compute the transition densities by simulation methods, and Aït-Sahalia [1, 2] provides series expansions of the transition densities. These approaches require the so-called reducibility condition [2, Definition 1] in the multivariate setting. Moreover, the number of terms needed in both simulation methods and series expansions to achieve a desired accuracy grows exponentially in the dimension of the SDE. As opposed to this we expect the computation of the Laplace approximation to be feasible for high dimensional SDEs. This postulate remains to be studied in future work.

**Example 1.** *The Brownian motion  $X$  with diffusion coefficient  $\theta$  is described by the flow  $\phi_x^{0,t}(f) = x + \theta f(t)$ . The Euler-Lagrange equation reads  $\partial_t^2 \widehat{X}(t) = 0$ . Thus, the MAP path from  $\widehat{X}(0) = x_0$  to  $\widehat{X}(T) = x_T$  is a straight line and*

$$\partial_t \widehat{B}(t) = \frac{x_T - x_0}{\theta T}, \quad \|\partial \widehat{B}\|_{\mathcal{K}}^2 = \frac{1}{2} \int_0^T \frac{(x_T - x_0)^2}{\theta^2 T^2} dt = \frac{(x_T - x_0)^2}{2T\theta^2}.$$

We have  $\alpha_f(t) = 0$  and  $\gamma_f(t) = \theta$ , and the Laplace approximation stated in Theorem 6 provides the exact transition density

$$p(x_T) = \frac{\exp\left(-\frac{1}{2}\|\partial\widehat{B}\|_{\mathcal{K}}^2\right)}{\sqrt{2\pi\int_0^T\gamma_f(t)^\top\gamma_f(t)dt}} = \frac{1}{\sqrt{2\pi T\theta^2}}\exp\left(-\frac{(x_T-x_0)^2}{2T\theta^2}\right).$$

The invariance property of the Laplace approximation may be illustrated considering the geometric Brownian motion  $\exp\left(X(t)-\frac{\theta^2}{2}t\right)$ . For the geometric Brownian motion we have  $\alpha_f(t) \neq 0$ , and the calculation of the Laplace approximation is more involved. However, the invariance property eq. (23) can be verified by numerical computations.

**Example 2.** The Ornstein-Uhlenbeck process  $dX(t) = -\theta X(t)dt + dB(t)$  with initial value  $X(0) = x_0$  and parameter  $\theta > 0$  is given by the flow

$$\phi_{x_0}^{0,t}(f) = e^{-\theta t}x_0 + \int_0^t e^{-\theta(t-s)}df(s).$$

The Euler-Lagrange equation reads  $\partial_t^2\widehat{X}(t) = \theta^2\widehat{X}(t)$ . Thus, the MAP Ornstein-Uhlenbeck path from  $\widehat{X}(0) = x_0$  to  $\widehat{X}(T) = x_T$  is

$$\widehat{X}(t) = \frac{(e^{\theta(T-t)} - e^{-\theta(T-t)})x_0 + (e^{\theta t} - e^{-\theta t})x_T}{e^{\theta T} - e^{-\theta T}}.$$

The relationship  $\partial_t\widehat{B}(t) = \partial_t\widehat{X}(t) + \theta\widehat{X}(t)$  gives

$$\|\partial\widehat{B}\|_{\mathcal{K}}^2 = \left(\frac{2\theta(x_T - e^{-\theta T}x_0)}{e^{\theta T} - e^{-\theta T}}\right)^2 \int_0^T e^{2\theta t}dt = \frac{2\theta(x_T - e^{-\theta T}x_0)^2}{1 - e^{-2\theta T}}.$$

We have  $\alpha_f(t) = 0$  and  $\gamma_f(t) = e^{-\theta(T-t)}$ , and the Laplace approximation stated in Theorem 6 provides the exact transition density

$$p(x_T) = \frac{\exp\left(-\frac{1}{2}\|\partial\widehat{B}\|_{\mathcal{K}}^2\right)}{\sqrt{2\pi\int_0^T\gamma_f(t)^\top\gamma_f(t)dt}} = \sqrt{\frac{\theta}{\pi(1 - e^{-2\theta T})}}\exp\left(-\frac{\theta(x_T - e^{-\theta T}x_0)^2}{1 - e^{-2\theta T}}\right).$$

**Example 3.** The squared Bessel process  $X(t)$  of dimension  $\theta > 1$  and the Euler-Lagrange equation for the associated MAP path  $\widehat{X}(t)$  are given by

$$dX(t) = (\theta - 1)dt + 2\sqrt{X(t)}\circ dB(t), \quad 2\widehat{X}(t)\partial_t^2\widehat{X}(t) = \widehat{X}(t)^2 - (\theta - 1)^2.$$

Strictly speaking the squared Bessel process is not covered by the developed theory: Assumption (A1) is violated since the diffusion coefficient  $\sigma(x) = 2\sqrt{x}$  lies in  $\mathcal{C}^{k+1,\eta}$  and not in  $\mathcal{C}_b^{k+1,\eta}$  [13, p. 334]. This is due to the behavior at  $x = 0$ . But since the neighborhood of  $x = 0$  is not entered by the MAP path we can still use the Laplace approximation. The squared Bessel process

has been chosen as an example for two reasons. Firstly, the Euler-Lagrange equation for the MAP path can be explicitly solved. We have  $\hat{X}(t) = c_2 t^2 + c_1 t + c_0$ . The coefficients  $c_0, c_1, c_2$  are found solving a quadratic equation in the boundary conditions and the MAP path from  $X(0) = x_0$  to  $X(T) = x_T$  is given by

$$\hat{X}(t) = \frac{T^2 - t^2}{T^2} x_0 + \frac{t^2}{T^2} x_T + \frac{t(T - t)}{T} c_1, \quad c_1 = \frac{-2x_0}{T} + \sqrt{\frac{4x_0 x_T}{T^2} + (\theta - 1)^2}.$$

Secondly, the transition densities and the optimal martingale estimator studied by Bibby & Sørensen [8] are explicitly known. This allow us to compare the approximate ML estimator based on the Laplace approximation with the exact ML estimator and the optimal martingale estimator. To do this we have minimized  $\mathbb{E}_{\theta_0}^{x_0}[-\log \hat{p}_{\theta}^{x_0}(X(1))]$  over  $\theta > 1$ . Here the expectation is taken over the squared Bessel process at time  $T = 1$  in dimension  $\theta_0 = 4$  conditioned on  $X(0) = x_0$ , and  $\hat{p}_{\theta}^{x_0}$  denotes the Laplace approximations of the transition densities in dimension  $\theta > 1$ . Since the diffusion coefficient  $2\sqrt{x}$  is more linear for larger values of  $x$  we expect the quality of the Laplace approximation to improve with larger values of  $x_0$ .

The approximate ML estimator as a function of  $x_0 \in (0, 25]$  is shown in the first panel in Figure 1. The approximate MLE is biased, but the bias is much smaller than the standard deviation expected from the Fisher information displayed in the upper right panel. Moreover, the Fisher information for the approximate and the exact ML estimators are almost identical and distinctively larger than the information of the optimal martingale estimator. The lower left panel in Figure 1 shows the entropy  $\min_{\theta > 1} \mathbb{E}_{\theta_0}^{x_0}[-\log \hat{p}_{\theta}^{x_0}(X(1))]$  as a function of  $x_0 \in (0, 25]$ . The difference between the approximate and the exact entropy can be decomposed in two parts. The first part,  $-\log \int \hat{p}_{\theta}^{x_0}(y) dy$ , arises since the Laplace approximation not necessarily provides a probability density. The second part is the Kullback-Leibler divergence between the normalized Laplace approximation and the exact transition density. In this case, the approximate and the exact entropy are almost identical. The last panel of Figure 1 displays the Laplace approximation of the transition density for  $x_0 = 1$ . Although this initial value resulted in a biased estimator, the approximation of the transition density is quite good. Furthermore, we see that  $-\log \int \hat{p}_{\theta}^1(y) dy > 0$ .

## A ODEs for the Laplace approximation

In this appendix we present the system of ordinary differential equations to be solved in the special case  $q = d = J = 1$ . Suppose we are given the data

$$x_0, \dot{x}_0, b(x), \partial_x b(x), \partial_x^2 b(x), \sigma(x), \partial_x \sigma(x), \partial_x^2 \sigma(x)$$

and a  $z \in [0, 1]$ . To find the energy and the trace terms needed in Theorem 6 we successively solve three subsystems S1, S2, S3 of the system of ODEs



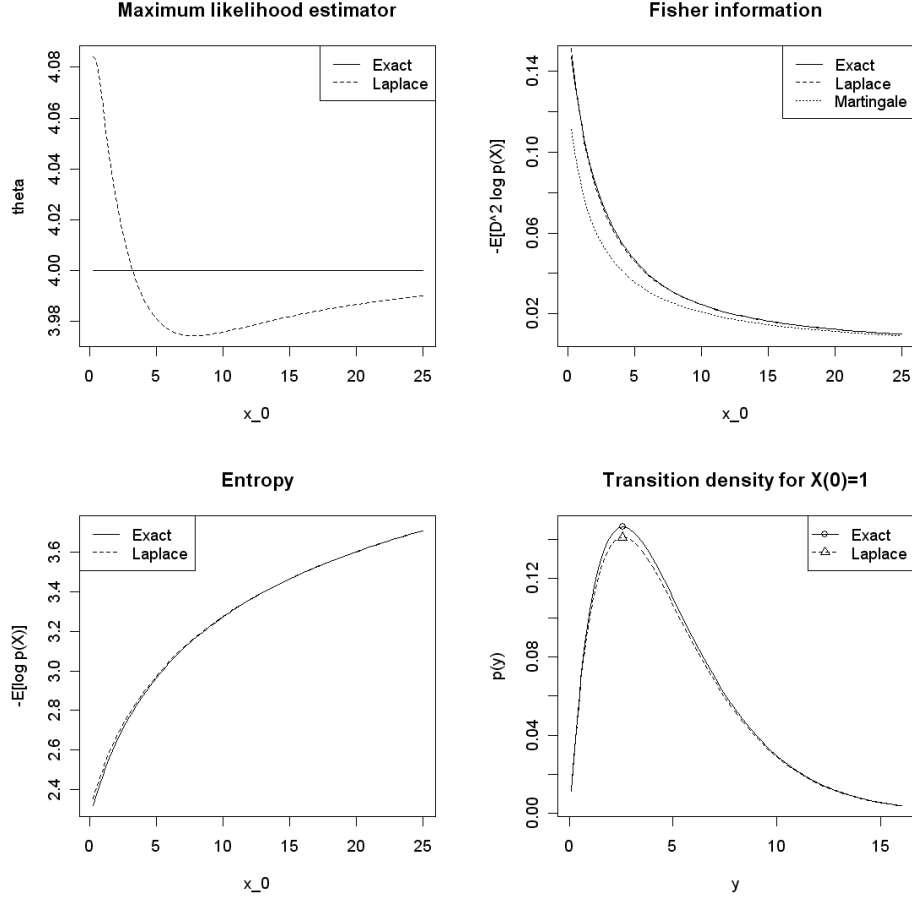


Figure 1: Comparison of the exact MLE, the approximate MLE and the optimal martingale estimator for the squared Bessel process. The first three panels show the estimates, their Fisher information and the entropy as a function of the initial position  $X(0) = x_0$ . The last panel shows the Laplace approximation of the transition density for  $x_0 = 1$ .

given by the initial values  $X(0) = x_0$ ,  $\partial_t X(t) = \dot{x}_0$  for  $t = 0$ ,  $\kappa(0) = 1$ ,  $\lambda(0) = 0$ ,  $\varphi(0) = 0$ ,  $\psi(0) = 1$ ,  $\rho_0(0) = 0$  or  $\rho(0) = -\rho_0(T)$ ,

$$\alpha(0) = -\frac{\partial_T B(T)}{\sigma(X(T))} \kappa(T) \left( \lambda(T) + \frac{\partial_x \sigma(x)}{\sigma(x)} \Big|_{x=X(0)} \right),$$

and derivatives

$$\begin{aligned}
\partial_t X(t) &= \partial_t X(t), \\
\partial_t^2 X(t) &= \frac{\partial_x \sigma(x)}{\sigma(x)} \left( (\partial_t X(t))^2 - b(x)^2 \right) + b(x) \partial_x b(x) \Big|_{x=X(t)}, \\
\partial_t \kappa(t) &= \kappa(t) \left( \partial_x b(x) + \partial_x \sigma(x) \partial_t B(t) \right) \Big|_{x=X(t)}, \\
\partial_t \lambda(t) &= \kappa(t) \left( \partial_x^2 b(x) + \partial_x^2 \sigma(x) \partial_t B(t) \right) \Big|_{x=X(t)}, \\
\partial_t \alpha(t) &= -\frac{\partial_T B(T)}{\sigma(X(T))} \kappa(T) (\partial_t \xi_t - \partial_t \lambda(t)), \\
\partial_t \varphi(t) &= \sigma(X(t))^2 \kappa(t)^{-2} + z \varphi(t) \partial_t \alpha(t) \varphi(t), \\
\partial_t \psi(t) &= z \varphi(t) \partial_t \alpha(t) \psi(t), \\
\partial_t \rho_0(t) &= \partial_t \rho(t) = -z \psi(t) \partial_t \alpha(t) \psi(t).
\end{aligned}$$

Here  $\partial_t B(t) = \frac{\partial_t X(t) - b(x)}{\sigma(x)} \Big|_{x=X(t)}$  and

$$\partial_t \xi(t) = \left( \partial_t \kappa(t) \frac{\partial_x \sigma(x)}{\sigma(x)} + \kappa(t) \frac{\partial_x^2 \sigma(x)}{\sigma(x)} \partial_t X(t) - \kappa(t) \frac{(\partial_x \sigma(x))^2}{\sigma(x)^2} \partial_t X(t) \right) \Big|_{x=X(t)}.$$

Furthermore, we define  $\Delta_T \alpha(T) = -\partial_T B(T) \kappa(T)^2 \partial_x \sigma(X(T)) \sigma(X(T))^{-2}$  and

$$\begin{aligned}
\varphi(T) &= (1 - z \varphi(T-) \Delta_T \alpha(T))^{-1} \varphi(T-), \\
\psi(T) &= (1 + z \Delta_T \alpha(T) \varphi(T)) \psi(T-), \\
\rho_0(T) &= \rho_0(T-) - z \psi(T-) \Delta_T \alpha(T) \psi(T)
\end{aligned}$$

The energy and the trace terms are given by  $\|\partial B\|_{\mathcal{K}}^2 = \text{Int}1(T)$  and

$$\begin{aligned}
\int_0^T \gamma(s)^\top \gamma(s) \, ds &= \kappa(T) \text{Int}2(T) \kappa(T), & \text{tr}[HA] &= \text{Int}5(T) + \text{Int}6(T), \\
GHG^\top &= \kappa(T) \text{Int}7(T), & GHAHG^\top &= \text{Int}8(T),
\end{aligned}$$

where the integrals Int1, ..., Int8 are found in the systems S1, S2, S3 as follows:

$$\begin{aligned}
S1 : \quad \text{Int1}(t) &= \int_0^t \partial_s B(s)^\top \partial_s B(s) \, ds, \\
S1 : \quad \text{Int2}(t) &= \int_0^t \beta(s)^\top \beta(s) \, ds, \\
S2, S3 : \quad \text{Int3}(t) &= \int_0^t \psi(s)^{-1, \top} \beta(s)^\top \beta(s) \, ds, \\
S2, S3 : \quad \text{Int4}(t) &= \int_0^t \rho_0(s) \psi(s)^{-1, \top} \beta(s)^\top \beta(s) \, ds - \rho_0(T) \text{Int3}(t), \\
S3 : \quad \text{Int5}(t) &= \int_0^t \alpha(s) \beta(s)^\top \beta(s) \, ds, \\
S3 : \quad \text{Int6}(t) &= 2 \int_0^t \alpha(s) \beta(s)^\top \beta(s) \psi(s)^{-1} \rho(s) \text{Int3}(s) \, ds, \\
S3 : \quad \text{Int7}(t) &= \int_0^t \beta(s)^\top H \gamma(s) \, ds, \\
S3 : \quad \text{Int8}(t) &= 2 \int_0^t H \gamma(s)^\top \beta(s) \alpha(s) \text{Int7}(s) \, ds.
\end{aligned}$$

Here  $\beta(t) = \sigma(X(t)) \kappa(t)^{-1}$ ,  $\rho(t) = \rho_0(t) - \rho_0(T)$  and

$$\begin{aligned}
H \gamma(t) &= \beta(t) \kappa(T) + \beta(t) \psi(t)^{-1} \rho(t) \text{Int3}(t) \kappa(T) \\
&\quad + \beta(t) \psi(t)^{-1} \text{Int4}(t) \kappa(T) - \beta(t) \psi(t)^{-1} \text{Int4}(t) \kappa(T).
\end{aligned}$$

**Acknowledgement.** I am grateful to Peter Johansen and Mads Nielsen for initiating my interest in Brownian flows. I wish to thank Martin Jacobsen for helping with the presentation and asking critical questions, and Helle Sørensen for reading the manuscript. Many thanks go to the anonymous referee who's careful and insightful reports resulted in substantial improvements of the paper.

## References

- [1] Y. Aït-Sahalia, “Maximum likelihood estimation of discretely sampled diffusions: a closed form approximation approach”, *Econometrica*, vol. 70, 223–262, 2002.
- [2] Y. Aït-Sahalia, “Closed-form likelihood expansions for multivariate diffusions”, working paper (available from <http://www.princeton.edu/~yacine/research.htm>), Princeton, 2004.

- [3] F. Baudoin, “Conditioned stochastic differential equations: theory, examples and application to finance”, *Stoch. Proc. Appl.*, vol. 100, 109–145, 2002.
- [4] J.O. Berger, *Statistical Decision Theory and Bayes Statistics*, Springer Verlag, 1999.
- [5] A. Beskos, O. Papaspiliopoulos, G.O. Roberts & P. Fearnhead, “Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes”, *J. R. Statist. Soc. B*, vol. 68, part 3, 333–382, 2006.
- [6] B.M. Bibby, M. Jacobsen & M. Sørensen, “Estimating functions for Discretely Sampled Diffusion-Type Models”. In *Handbook of Financial Econometrics* (eds. Y. Ait-sahalia & L.P. Hansen), North-Holland, 2004.
- [7] F.H. Clarke, “Tonelli’s regularity theory in the calculus of variations: Recent progress”, *Lecture Notes in Mathematics*, vol. 1190, 163–179, 1986.
- [8] B.M. Bibby & M. Sørensen, “Martingale estimation functions for discretely observed diffusion processes”, *Bernoulli*, vol. 1, 17–39, 1995.
- [9] D. de Falco & D.C. Khandekar, “Application of white noise calculus to the computation of Feynman integrals”, *Stoch. Proc. Appl.*, vol. 29, 257–266, 1988.
- [10] R.D. Gill & S. Johansen, “A survey of product-integration with a view toward application in survival analysis”, *Ann. Statist.*, vol. 18, 1501–1555, 1990.
- [11] S.L. Heston, “A closed-form solution for options with stochastic volatility with applications to bond and currency options”, *Rev. Financial Studies*, vol. 6, 327–343, 1993.
- [12] T. Hida, H.-H. Kuo, J. Potthoff & L. Streit, *White noise*, Kluwer, 1993.
- [13] H. Kunita, *Stochastic Flows and Stochastic Differential Equations*, Cambridge University Press, 1990.
- [14] B. Markussen, “A statistical approach to large deformation diffeomorphisms”, workshop on Generative Model Based Vision at the IEEE Conference on Computer Vision and Pattern Recognition, Washington DC, 2004.
- [15] B. Markussen, “Large deformation diffeomorphisms with application to optic flow”, *Computer Vision and Image Understanding*, vol. 106, 97–105, 2007.

- [16] B. Markussen, “Inversion of some integral operators with a lattice structure”, working paper (<http://www.matfys.kvl.dk/~torbenm/publ.dept/2008.html>), Department of Natural Sciences, University of Copenhagen, 2008.
- [17] D. Nualart, *Malliavin Calculus and Related Topics*, Springer Verlag, 1995.
- [18] G.K. Pedersen, *Analysis Now*, Springer Verlag, 1989.
- [19] R. Poulsen, “Approximate maximum likelihood estimation of discretely observed diffusion processes”, working paper no. 29, Center of Analytic Finance, Aarhus, 1999.
- [20] H. Sørensen, “Parametric inference for diffusion processes observed at discrete points in time: a survey”, *Int. Statist. Rev.*, vol. 72, 337–354, 2004.
- [21] L. Streit & T. Hida, “Generalized Brownian functionals and the Feynman integral”, *Stoch. Proc. Appl.*, vol. 16, 55–69, 1984.
- [22] L. Tonelli, “Sur une méthode directe du calcul des variations”, *Rend. Circ.*, vol. 39, 233–264, 1915.
- [23] C. Twining & S. Marsland, “Constructing diffeomorphic representations of non-rigid registrations of medical images”, *Lecture Notes in Computer Science*, vol. 2732, 413–425, 2003.